The Molecular Genetics Of Polyketide Biosynthesis In Filamentous Fungi

Lewis Edward Hector Bingle

A thesis submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science, School of Biological Sciences

December 1997

ABSTRACT

A genomic library was screened to determine whether genes in the biosynthetic pathway leading to the polyketide secondary metabolite patulin are clustered in the filamentous fungus *Penicillium patulum.* The regions of chromosome around the first gene in the pathway, which encodes 6-methylsalicylic acid synthase (MSAS), were examined for the presence of co-expressed linked genes. By means of chromosome walking, DNA sequence analysis and northern blotting two open reading frames (ORFs) were identified in the 6 kb upstream of *MSAS* that corresponded to concurrently transcribed regions of chromosome. These were designated ORF1, the source of a 1.5 kb transcript and ORF2, the source of a 2 kb transcript. Comparison of the predicted gene product for ORF2 with a protein sequence database indicated high levels of similarity to cytochrome P450 enzymes, a type of enzyme thought to be involved in the patulin biosynthetic pathway. Four introns were identified in the *ORF2* gene ranging in size from 50 to 105 bp. To facilitate assessment of the enzymic activity of ORF2, a full length cDNA copy was expressed in *E. coli* as a fusion to the carboxy-terminus of glutathione S-transferase.

Two pairs of degenerate PCR primers (LC1 and LC2c, LC3 and LC5c) were designed for the amplification of ketosynthase domain fragments from novel fungal PKS genes. Both primer pairs were shown to amplify one or more PCR products from a range of deuteromycete genomes. Southern blotting analysis confirmed that the products obtained from a genome using each pair of primers were amplified from structurally distinct loci. PCR products obtained with the LC1/2c primer pair from *Penicillium patulum* and *Aspergillus parasiticus* were cloned and sequenced; a *Verticillium dahliae* PCR product was sequenced directly. Comparison of the sequenced PCR products with protein sequences in the SwissProt database showed that the polypeptides encoded by the amplified DNA fragments were highly homologous to other type I PKS genes. A simple phylogenetic analysis suggested that the genes from which these fragments were amplified were closely related to fungal PKS genes involved in pigment and aflatoxin biosynthetic pathways.

A 6.5 kb fragment of *Penicillium patulum* genomic DNA containing a complete novel PKS gene (*PKS2*)was sequenced. Analysis of the *PKS2* sequence suggested that the predicted gene product encoded ketosynthase, acyltransferase, dehydratase, ketoreductase and acyl carrier protein domains and was structurally similar to MSAS. Northern analysis showed that *PKS2* is not expressed under culture conditions where *MSAS* is transcribed. A closely linked and divergently transcribed cytochrome P450 gene was identified in the region of chromosome downstream of *PKS2*.

Acknowledgements

I am very grateful for the excellent supervision and guidance of Dr Colin Lazarus throughout this project.

I thank my tutor Professor Tom Simpson for his advice and encouragement.

The support of my family, in particular my mother and father, is acknowledged with deep gratitude.

Special thanks to Lucy for all her support during the latter stages of my research and the writing of this thesis.

I am grateful to the following people: Dr Steve Screen for large amounts of advice and the gift of plasmid pGEX-BHE(E)XSX, Dr John Crosby for assays of enzymic function, Dr Gary Barker and Dr Jamie Stevens for help and advice on phylogenetic matters, Dr Heather Macdonald for advice in the lab, Professor G. Maurice Gaucher, Robin Couch and Sunil Shah for information and advice on matters relating to patulin biosynthesis and Dr Andy Bailey for donating the cytochrome c degenerate PCR primers.

Thank you to all workers in BC9, past and present, for making my time here so enjoyable.

A grant from the BBSRC is gratefully acknowledged.

Author's declaration

The research presented in this thesis is the result of my own independent work carried out in the School of Biological Sciences under the supervision of Dr Colin Lazarus, unless explicitly stated otherwise. The views expressed herein are those of the author and not of the University of Bristol.

22nd December 1997

Contents

Figures and Tables

Chapter 1

Introduction

1.1 POLYKETIDES

1.1.1 Origins of the polyketide hypothesis

The biosynthetic origins of the polyketide metabolites were established by the sequential contributions of three workers, John N Collie, Robert Robinson and Arthur J Birch (Birch, 1967; O'Hagan, 1991). Collie proposed that ketene (CH_2=C=O) might be the basic unit from which a number of aromatic plant metabolites are constructed and termed these metabolites *polyketides* (Collie, 1907). However this analysis had little impact on contemporary scientists and his basic hypothesis was not built on for four decades. By this time the involvement of acetate in fatty acid biosynthesis had been established by feeding ^{13}C- and deuterium-labelled acetates to mice and rats then measuring the accumulation of these isotopes in their fatty acids (Rittenberg & Bloch, 1944, 1945). Sir Robert Robinson, inspired by this revelation, enlarged upon Collie's original hypothesis, predicting the origins of the tetracycline antibiotics along with other polyketides. Birch independently began to outline his "acetate hypothesis" in the early 1950s , stating that acetate was the building block of many aromatic plant metabolites (Birch & Donovan, 1953). Birch was influenced by Collie's hypothesis but also by contemporary research into fatty acid biosynthesis, which indicated that acetyl coenzyme A acts as an "active acetate". This version of the hypothesis was supported by extensive experimentation, both indirectly by its use in structural analysis of aromatic polyketide metabolites (Birch, 1967) and directly by feeding experiments in *Penicillium griseofulvum,* wherein isotopically labelled acetate was shown to be incorporated into the polyketides 6-methylsalicylic acid (6MSA) and griseofulvin (Birch *et al.,* 1958, Birch, 1967).

Early research into polyketide biosynthesis tended to concentrate on the elucidation of biosynthetic pathways via identification of intermediate metabolites. More recent research efforts have tended to focus on the isolation and characterisation of pathway enzymes and enzyme genes and also on the biochemical and genetic mechanisms regulating the production of these compounds.

1.1.2 Occurrence in nature and biological activity

To quote David O'Hagan (1995) "There is no other class of natural products.....so extensively distributed in the biosphere". Polyketides probably represent the largest class of secondary metabolites and are produced by many organisms from bacteria to higher plants.

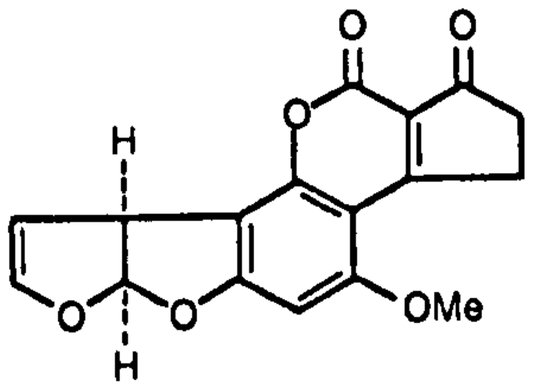
Some polyketide metabolites from various taxa are described below, with notes on their biological activity (reviewed by Hopwood & Sherman, 1990; O'Hagan, 1991,1995; Turner & Aldridge, 1983).

All plants produce flavonoids, such as the anthocyanins, which are responsible for flavour in many foods of plant origin and colouring in many flowers. Other flavonoids have a protective role, acting as UV-screening and antifeedant agents. The urishiols are irritants found in poison ivy. Stilbene and isoflavonoid phytoalexins are antifungal compounds produced in response to fungal infection. Various compounds with potential anticancer properties have been isolated from higher plants such as giganin, found in the bark of *Goniothalamus giganteus*.

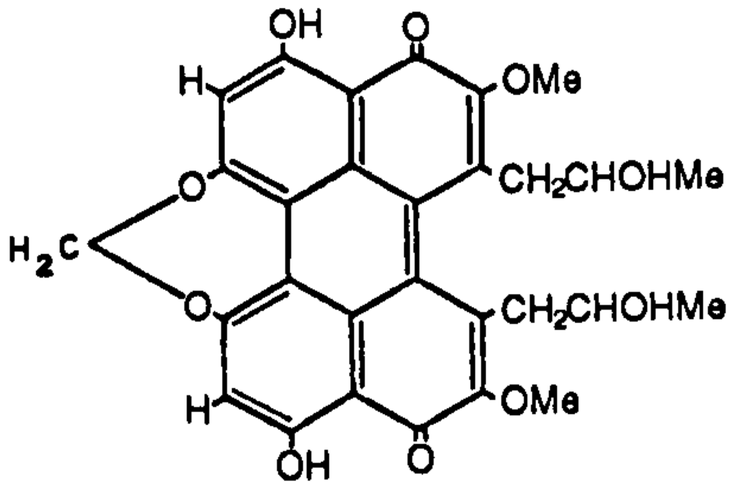
The fungal kingdom is the source of most of the polyketide secondary metabolites identified so far; some examples are shown in figure 1. Filamentous fungi often produce mycotoxins of polyketide origin (Steyn, 1992), including the aflatoxins and ochratoxins from *Aspergillus* spp., ergochromes from *Claviceps* and patulin from *Penicillium*. Several fungal polyketides have been identified as antibiotics, such as the antifungal griseofulvin and the antibacterial brefeldin A from *Penicillium* spp. Polyketide-derived dihydroxynapthalene melanins are required for infection by phytopathogenic fungi and some phytotoxins have a polyketide origin. Many of the vivid pigments which colour macromycete fruit bodies are polyketides, for example the anticancer agent austrocortirubin which is produced by the toadstool *Dermocybe splendida* (Gill, 1994). Compactin and mevinolin (also known as lovastatin), metabolites of *Aspergillus terreus*, are *in vivo* inhibitors of cholesterol biosynthesis and of obvious pharmaceutical interest.

The marine environment has proven to be a rich source of novel polyketides, which are produced by a diverse range of organisms. Examples include very elaborate dinoflagellate toxins like maitotoxin from *Gamberdiscus toxicus* and the brevetoxins produced by *Gymnodium* spp., also polypropionates similar to those produced by bacteria which are found in some species of molluscs.

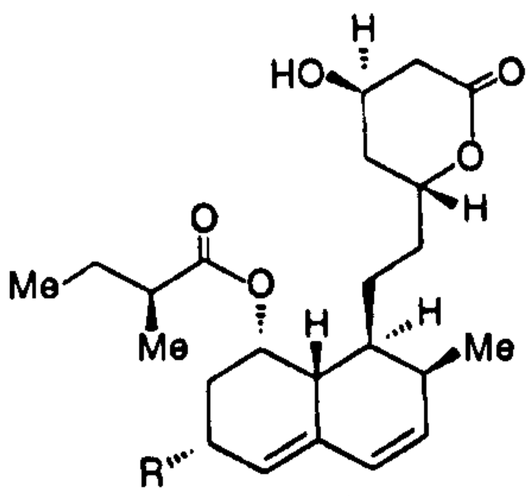
While insects are not usually considered as major producers of polyketides, the limited evidence available suggests that they do make use of polyketide biosynthetic pathways. Volatile aromatic polyketides such as methyl 6-methylsalicylic acid and mellein are used by various ant species as pheromones, playing a role in communication or signalling between individuals.



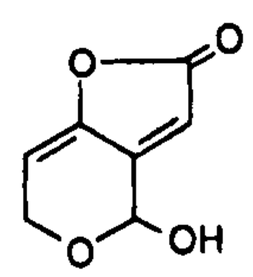
Aflatoxin B_1
mycotoxin



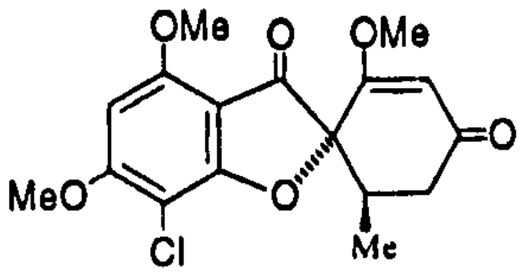
Cercosporin
phytotoxin



R = H, Compactin
R = Me, Mevinolin
inhibitors of cholesterol biosynthesis



Patulin
mycotoxin



Griseofulvin
antifungal

Figure 1.1 Examples of fungal polyketide structures and biological activities.

The most prolific prokaryotic producers of polyketides are the filamentous actinomycetes and the genus *Streptomyces* in particular. These bacteria elaborate hundreds of diverse antibiotics, including the tetracyclines, the anthracyclines, the macrolides (e.g. erythromycin), polyethers (e.g. monensin), ansamycins (e.g. rifamycin) and avermectin (an anti-helminthic). *Streptomyces* spp. also produce pharmacologically active compounds such as the immunosupressant macrolide FK506 and the antitumour drug aclacinomycin A. Some actinomycete polyketides may function as signal molecules in development (e.g. pamamycin). *Mycobacterium phlei* produces 6-methylsalicylic acid which is involved in iron uptake mechanisms.

A few polyketide antibiotics are produced by non-actinomycete bacteria, e.g. pseudomonic acid (mupirocin) from *Pseudomonas* and aurantinin from *Bacillus*. The Myxobacteria produce several diverse antibiotics. The plant pathogen *Pseudomonas syringae* produces the phytotoxin coronatine which has a polyketide component. Specificity signalling molecules for legume recognition produced by *Rhizobium leguminosum* and *R. meliloti* are thought to be polyketide in origin.

While polyketide metabolites seem to be most abundant in the actinomycete bacteria, the filamentous (lower) fungi, marine organisms and the higher plants, this could merely represent the intensity of interest in the polyketide metabolites of these groups, a consequence of their biological activity (Hopwood & Sherman 1990).

1.2 POLYKETIDES AS SECONDARY METABOLITES

1.2.1 Secondary metabolism

The biochemical processes in a living cell can be divided into two categories. Primary metabolism provides energy and raw materials for normal cell function and replicatory growth. Examples of primary metabolic processes include fatty acid biosynthesis, DNA replication, protein synthesis and glycolysis. Several succinct definitions of secondary metabolism have been proposed over the years (a selection is reviewed by Bennett, 1995); a useful contemporary one is: "biochemical pathways which are not necessary for the growth or reproduction of an organism, but which can be demonstrated genetically, physiologically or biochemically" (Hunter, 1992). Secondary metabolites, also referred to by some authors as natural products or idiolites, are of relatively low molecular weight and are very diverse in their chemistry. Due to their fundamental nature the processes of

primary metabolism are widely distributed in nature. This is in contrast to secondary metabolic pathways, which tend to have a restricted taxonomic distribution and are often specific to a particular genus, species or strain (Herbert, 1989). It has been suggested that the ability to produce secondary metabolites is associated with occupancy of particular ecological niches rather than with phylogeny (Hunter, 1992). An example can be seen in the filamentous fungi and the actinomycete bacteria: these taxonomically very distant groups tend to occupy the same ecological niche and both groups are prolific producers of secondary metabolites, while many species which are more closely related to either the bacteria or the fungi do not synthesise these products. Secondary metabolites are mainly synthesised from a small group of primary metabolites: acetyl-CoA, amino acids, mevalonic acid and the intermediates of the shikimic acid pathway (Herbert, 1989). The polyketides formed from acetyl-CoA probably represent the largest family of secondary metabolites; other families include the terpenes, steroids, alkaloids and shikimate metabolites.

While the importance of secondary metabolites to their producers is somewhat obscure, the relevance of these natural products to humans is beyond question. Secondary metabolites have provided a vast array of antibiotics and other pharmaceutically active agents, while fungal toxins have a huge impact on agricultural economics and human/livestock health. The active constituents of many recreational drugs of natural origin and also many perfumes and spices are all derived from secondary metabolism.

1.2.2 The evolution of secondary metabolism

Secondary metabolites, by definition, do not have an explicit role in the internal economy of the producer organism. Many theories, often conflicting, have been put forward to suggest ecological roles for these products and explain the evolution of secondary metabolism (reviewed by Haslam, 1986; Williams *et al.*, 1989; Stone & Williams, 1992). The enzymes involved in secondary metabolism are similar to their counterparts in primary metabolism and probably evolved by gene duplication and divergence from these counterparts (Hunter, 1992; Vining, 1992). Nucleotide and predicted amino acid sequences of secondary metabolic pathway genes often show much weaker homology to genes for related enzymes of primary metabolism in the same organism than to those of primary and secondary metabolism in other organisms. This suggests that horizontal gene transfer has been an important factor in the spread of secondary metabolism (Vining, 1992). It seems

unlikely that these complex and energetically expensive pathways would have evolved if they did not afford some selective advantage to the producer. A convincing argument has been made that secondary metabolites serve to increase the survival fitness of the producing organisms by their interaction with specific receptors in competing organisms (Stone & William, 1992; Williams *et al.*, 1989). Thus secondary metabolism may act as an alternative defence mechanism in organisms which lack an immune system. Analysis of the genetics of secondary metabolic biosynthetic pathways has provided some evidence in favour of this hypothesis, as described below (Stone & Williams, 1992).

The genes for secondary metabolite biosynthesis in both actinomycete bacteria and filamentous fungi are generally clustered (Bennett, 1995; Hopwood & Sherman, 1990; Keller *et al.*, 1992; Keller & Hohn, 1997; Smith *et al.*, 1990). This linkage of pathway genes implies that they have evolved as a unit to some extent and increases the probability that they will be passed on as a unit to subsequent generations or other species. Clustering of these genes would only be selected for if the secondary metabolite produced conferred some selective advantage on the producing organism. The fact that these clusters contain only functional genes involved in the biosynthetic pathway suggests that they could not have arisen by a series of evolutionarily neutral gene duplications and mutations.

Many *Streptomyces* secondary metabolites are antibiotics. In some cases these antibiotics would kill the producing organism were it not for the action of resistance genes, which are often closely linked to the antibiotic biosynthesis genes . This linkage can be explained as a consequence of the lateral transfer of antibiotic biosynthesis genes between species; clustering of these genes improves the chances of the recipient organism receiving both antibiotic biosynthetic and resistance genes, hence avoiding self-destruction (Hopwood, 1988). The existence of antibiotic resistance genes and their clustering with biosynthetic genes in this manner provides evidence that the antibiotics are produced by these organisms in their natural habitat and do exert some selective pressure.

Links between sporulation and secondary metabolism have been reported in the bacilli and in various actinomycete species (reviewed by Betina, 1995). Genetic analysis of the actinomycete model system, *Streptomyces coelicolor* A3(2), has identified several genes which are essential for both aerial mycelium formation (a prerequisite for sporulation) and secondary metabolism (Hopwood, 1988). The dual role identified for these genes, known as pleiotropic switches, clearly indicates that aerial mycelium formation and secondary metabolism are interrelated. The predicted products of genes involved in spore pigment

synthesis in both actinomycetes and fungi have been shown to have a high level of sequence homology to genes involved in polyketide secondary metabolite biosynthesis (Davis & Chater, 1990; Mayorga & Timberlake, 1992). This may indicate that antibiotic biosynthesis genes evolved from genes with a functional role in the sporulation process, with both sets of genes remaining under a common genetic control mechanism. Alternatively pleitropic switching may have evolved because the switching on of secondary metabolism will provide a defense mechanism and increase competitiveness during aerial mycelium formation and sporulation, processes which use up a great deal of metabolic energy and may leave the producing organism vulnerable. Pleiotropic switching may also be seen as a two-pronged defensive response to low nutrient levels and the accompanying increase in competition.

Many antibiotic secondary metabolites interact with ribosomal and other RNA molecules, affecting intron splicing, peptide bond formation and other processes involved in protein biosynthesis (Davies, 1990; Hunter, 1992). A putative primitive function as effectors of protein synthesis could have evolved into the antagonistic role of today's antibiotics, many of which act on the ribosome (Davies, 1990; Hunter, 1992). This suggestion could imply a broader role for the primordial predecessors of the secondary metabolites as enzyme effectors or catalytic agents in various biosynthetic reactions. Another suggested primordial role for secondary metabolites is as agents of cellular communication; many are known to be 'receptor active' and some microbes have been shown to have vertebrate-like receptors (Hunter, 1992; Roth *et al.*, 1986).

1.3 POLYKETIDE BIOSYNTHESIS

1.3.1 Polyketide chain assembly

The polyketide metabolites exhibit a huge range of structural diversity, from simple aromatic compounds such as 6-methylsalicylic acid to gigantic, complex molecules like the polycyclic ether maitotoxin, the largest known secondary metabolite with a molecular weight of 3.4 kDa (O'Hagan, 1995). In general, fungal and plant polyketides tend to be aromatic and structurally less complex than those of bacteria and marine organisms. However they are all formed by the decarboxylative condensation of small carboxylic acids in a head to tail fashion to form a carbon "backbone" (figure 2). Each carboxylate unit contributes two carbon atoms to the growing chain, the ß-carbon carrying a keto group.

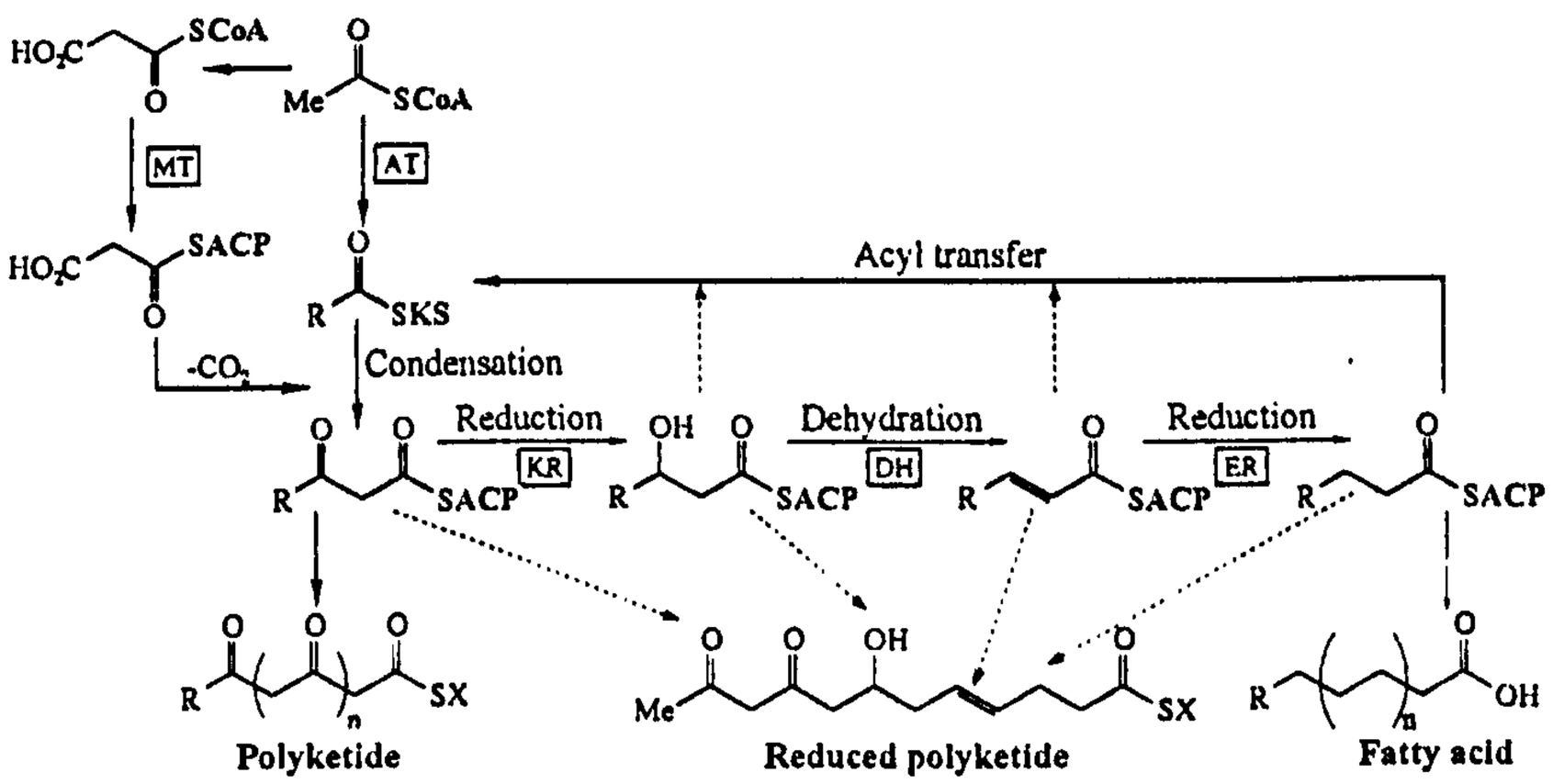


Figure 1.2 Polyketide biosynthesis.

Adapted from Simpson (1995).

KS, ketosynthase; AT, acyltransferase; MT, malonyltransferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase ACP, acyl carrier protein; CoA, Co-enzyme A.

According to the number of two-carbon units that have contributed to its biosynthesis, the resultant chain is designated a tetraketide, pentaketide, etc. Each condensation may be followed by a cycle of modifying reactions on the keto group, including some or all of the following: ketoreduction, dehydration and enoyl reduction. The keto groups on the carbon backbone may undergo all these reactions and be completely reduced to methylene groups as in the fatty acids. The other extreme is for no reduction to take place resulting in an extended poly-ß-keto chain. Many polyketide metabolites show a degree of reduction somewhere between these two extremes, the result of selective deployment of the reducing and dehydrating activities.

The "starter" unit for the carbon chain may be an acyl thioester such as acetyl-CoA, propionyl-CoA, benzoyl-CoA, or a more complex thioester of e.g. cinnamate or an amino acid. Extension of the carbon chain unit will be by malonyl-CoA units (often referred to as "acetate"). In bacterial or marine polyketides methylmalonyl-CoA ("propionate") or ethylmalonyl-CoA ("butyrate") may extend the growing chain, giving rise to pendant methyl or ethyl groups. While fungi are not known to incorporate extender units other than malonate, additional branching methyl groups may be seen in some polyketide metabolites of fungi or plants; these are derived from L-methionine (Birch, 1958).

The huge range of naturally-occurring polyketide structure can be attributed to the following features of polyketide biosynthesis (O'Hagan, 1991; Simpson, 1995):

1. Selective processing of keto groups.
2. Use of a variety of chain starter and extender units.
3. Variation in overall chain length and cyclisation of the polyketide "backbone".
4. Subsequent modifications of the polyketide backbone.

1.3.2 Fatty acid synthase enzymes

Fatty acids are primary metabolites, found in all living cells due to their roles as membrane components and as an energy store in the form of triglycerides. Fatty acid biosynthesis (reviewed by O'Hagan, 1991, 1995) is closely analogous to polyketide biosynthesis, requiring similar enzymatic activities. Indeed fatty acid biosynthesis may be considered as a subset of polyketide biosynthesis in which a restricted range of substrates is utilised, all keto groups are completely reduced to methylene and chain termination is usually accompanied by further processing of the carbon chain. Until the discovery of the modular PKS enzymes responsible for complex polyketide assembly (to be discussed below) it was

believed that all polyketides were assembled by synthases related to fatty acid synthases (FAS), the enzymes responsible for long chain fatty acid (LCFA) biosynthesis. All FAS systems fall into two distinct classes: type I systems consist of a multifunctional polypeptide possessing the required catalytic domains for LCFA biosynthesis and are found in some actinomycetes, fungi and animal systems; the type II FAS, found in bacteria such as *E. coli*, cyanobacteria and plants comprises a number of separate polypeptides, each carrying a single catalytic domain.

Yeast and other fungal FAS enzymes consist of two multifunctional subunits, α and β, which between them comprise the full complement of enzymic activities for LCFA biosynthesis. The α subunit provides ketoreductase, ketosynthase and ACP domains and the ß subunit provides acetyl, malonyl and palmitoyl transferases, enoylreductase and dehydratase domains. The subunits each have a molecular mass of about 200 kDa and are arranged in a complex as an $\alpha_6\beta_6$ hexamer. FAS genes for the subunit proteins have been cloned from yeasts such as *Saccharomyces cerevisiae* and *Yarrowia lipolytica* (Köttig *et al.*, 1991; Mohamed *et al.*, 1988; Schweizer *et al.*, 1986) and the filamentous fungus *Penicillium patulum* (Wiesner *et al.*, 1988). The sequential order of catalytic domains appears to be identical in each case. There is no thioesterase domain apparent in these deduced amino acid sequences; the product of these FAS enzymes is a CoA thioester rather than the free fatty acid (Robinson, 1991).

The vertebrate type I FAS complex has a completely different architecture: all functional domains are found on a single polypeptide. *In vivo* the enzyme is found as a dimer with a typical native molecular mass of 540 kDa. Inhibition and cross-linking studies (reviewed by Alberts & Greenspan, 1984) have shown that the ketosynthase and acyl carrier protein thiol (-SH) groups, implicated in covalent binding of the growing carbon chain, are juxtaposed in the dimer, allowing them to interact during the condensation process. Genes for vertebrate FAS have been cloned from many species, including chicken (Holzer *et al.*, 1989), rat (Amy *et al.*, 1989; Schweizer *et al.*, 1989) and human (Jayakumar *et al.*,1995); all show a high colinear similarity over the various domains.

The type II FAS found in most bacteria consists of a minimum of seven discrete monofunctional enzymes and an ACP (O'Hagan 1991). There is no evidence to suggest that the discrete proteins of the type II FAS complex are aggregated within living cells and indeed aggregation is not necessary for activity *in vitro* (Robinson 1991). The bacterial type II system studied in greatest detail is palmitate synthase from *E. coli* (reviewed by

Hopwood & Sherman, 1991), which has three distinct ketosynthase enzymes. ß-Ketoacyl synthase III catalyses the condensation of malonyl-ACP with acetyl-CoA, rather than the usual substrate of acetyl-ACP (Tsay *et al.*, 1992). Some actinomycete bacteria such as *Mycobacterium tuberculosis* have a type I FAS which resembles a head to tail fusion of the two fungal subunits (Fernandes & Kolattukudy, 1996).

Plant FAS are generally associated exclusively with chloroplasts and are of the primitive bacterial type II category, consistent with the endosymbiotic origins of the chloroplasts. Multiple ketosynthase and ACP enzymes have been observed in plants (Hopwood & Sherman, 1990). FAS component genes have been cloned from a number of higher plant species (Töpfer & Martini, 1994). ACP has been studied most intensively and appears to be encoded by gene families (Töpfer & Martini, 1994).

1.3.3 Polyketide synthase enzymes

The biosynthesis of a polyketide metabolite is mediated by a multifunctional enzyme complex known as a polyketide synthase or PKS. In common with the fatty acid synthases these have been classified into two categories, the type I PKS which are single multifunctional polypeptides and the type II PKS which consist of a multi-enzyme complex wherein each component enzyme carries out a single catalytic function. A third category of PKS found in the actinomycete bacteria has been shown to have a modular organisation, described below.

A polyketide synthase will require several enzymic domains: a ketosynthase (condensing) domain; an acyl carrier protein (ACP), which binds the acylthioester intermediate at most stages in synthesis; acyl and malonyl transferases to load substrates onto the enzyme and a thioesterase activity to remove the finished product. In addition it may possess ketoreductase, dehydratase and enoylreductase functional domains for the processing of keto groups. All of these catalytic activities are structurally and functionally similar to the corresponding components of the FAS complex. A cyclase activity will also be present in PKS responsible for the synthesis of aromatic polyketides.

While polyketide metabolites seem to be widespread in nature, the enzymology of their biosynthesis has, so far, only been studied in bacteria, fungi and higher plants. Most efforts to purify active PKS for characterisation and mechanistic studies have been unsuccessful because of the size, complexity and instability of these enzymes.

1.3.3.1 Bacterial polyketide synthase enzymes

There has been no reported attempt to elucidate the biochemistry of a bacterial PKS enzyme by *in vitro* study of purified enzymes (Hutchinson & Fujii, 1995). Mechanistic studies have so far taken place *in vivo*, using recombinant bacteria to express native or mutant PKS genes or to examine the effect of mutations on the chromosomal PKS genes. The insights gained from these studies tend to be inferential. Reasons for this lack of progress include the low levels of expression of enzymes of secondary metabolism and the consequential difficulty of obtaining a purified enzyme preparation, as well as problems in devising useful assays based on cell-free activity (Robinson, 1991; Dutton, 1988). Luckily the genes encoding these enzymes have not been so difficult to isolate and gene sequences obtained have provided information on the primary structure of many bacterial PKS proteins. *In vivo* and cell-free activity of recombinant PKS from the erythromycin A and tetracenomycin biosynthetic pathways have been reported, and the availability of *in vitro* PKS systems should allow a direct investigation of the mechanisms of enzyme action (Kao *et al.*, 1994; Pieper *et al.*, 1995a,b; Shen & Hutchinson, 1993b). Researchers in Professor T.J. Simpson's group at the University of Bristol have succeeded in expressing several type II *Streptomyces* PKS acyl carrier proteins in *E. coli*. High yields of expressed protein have been achieved, with the active *holo*-ACP being produced in some cases (Cox *et al.*, 1997; Crosby *et al.*, 1995).Heterologous expression of these ACPs will facilitate the study of ACP structure and function and has already led to the determination of the solution structure of the *act apo*-ACP, the only three-dimensional structure of a PKS component elucidated to date (Crump *et al.*, 1997).

1.3.3.2 Fungal polyketide synthase enzymes

From the examples described to date, mostly predicted from gene sequences, it appears that fungal PKS enzymes fall into the type I category. All the catalytic domains required for synthesis of the polyketide product are found on a single large polypeptide, encoded by a single gene.

Studies of enzymology carried out on fungal PKS systems are described in detail below.

1.3.3.2.1 6-Methylsalicylic acid synthase

The fungal PKS which has to been studied most extensively is 6-methylsalicylic acid synthase (6MSAS) from *Penicillium patulum.* This enzyme catalyses the formation of the

tetraketide 6-methylsalicylic acid (6MSA), the first stable free intermediate during biosynthesis of the mycotoxin patulin, from one molecule of acetyl-CoA and three molecules of malonyl-CoA (figure 3).

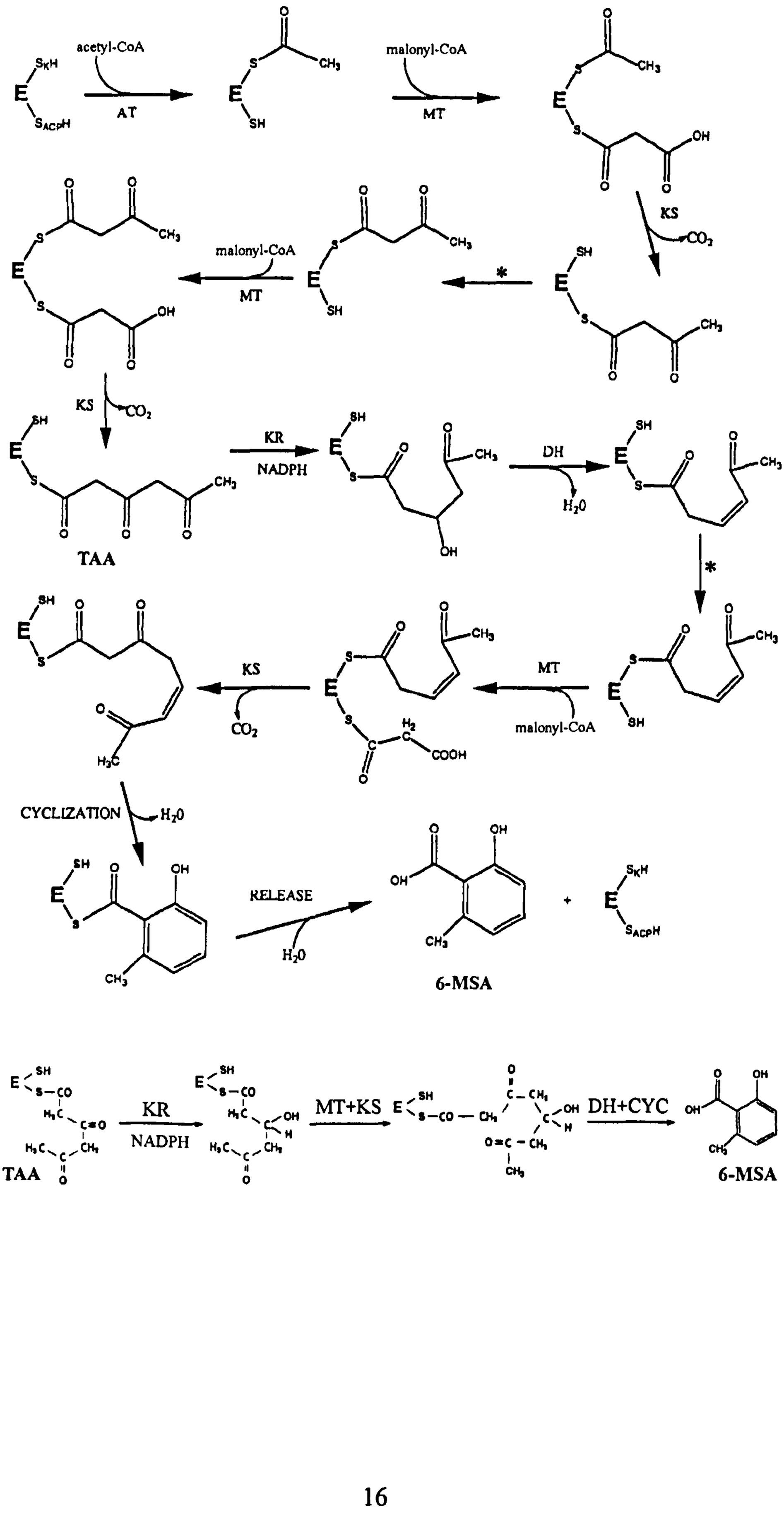
This enzyme has been purified to varying extents by several groups (Dimroth *et al.*, 1970,1976; Scott *et al.* 1974; Spencer & Jordan, 1992a). 6MSAS is highly susceptible to *in vivo* and *in vitro* proteolytic degradation, but may be stabilised by incorporating proteinase inhibitors into the purification buffers (Spencer & Jordan, 1992a). The substrate specificity of 6MSAS has been assayed and K_m values obtained for acetyl- and malonyl-CoA in the range 7-20 μM (Dimroth *et al.*, 1970; Spencer & Jordan, 1992a). The enzyme has been shown to accept propionyl-CoA in place of acetyl-CoA as a chain starter unit (Dimroth *et al.*, 1976). In the presence of propionyl-CoA, synthesis of 6-ethylsalicylic acid was found to occur at 13% of the normal rate for 6-methylsalicylic acid. An acetyl transferase activity, analogous to that found in FAS, has been shown to transfer acetyl residues from acetyl-CoA to pantetheine and is also capable of catalysing propionyl transfer at a reduced rate. This reduced rate of synthesis with alternative starter units, together with a lower cellular concentration relative to acetyl-CoA, is thought to ensure that only 6MSA will be synthesised *in vivo*. Acetoacetyl-CoA, a partially assembled diketide intermediate, was also accepted as a substrate for chain elongation but it again resulted in a greatly reduced reaction rate, less than 5% of the normal value (Spencer & Jordan, 1992a). SDS-PAGE and gel filtration analysis characterised the native enzyme as a homotetramer with a subunit of 180 kDa (Spencer & Jordan, 1992a). Cross-linking studies using the bifunctional sulfhydryl reagent 1,3-dibromopropan-2-one indicated that the functional enzyme may take the form of a dimer with juxtaposed ketosynthase and ACP domains from two associated proteins making up the condensing active site, a similar arrangement to that found in the type I vertebrate FAS (Child *et al.*, 1996; Jordan & Spencer, 1993; Spencer & Jordan, 1992). *In vitro*, 6MSAS will undergo a chain terminating abortion reaction leading to the formation of the triketide triacetic acid lactone, which indicates that product release can occur before the normal chain length is reached (Schorr *et al.*, 1994). In the absence of the coenzyme NADPH triacetic acid lactone is formed as the exclusive product of 6MSAS (Dimroth *et al.*, 1970; Scott *et al.*, 1974; Spencer & Jordan, 1992a). This suggests that the enzyme-bound triketide is unable to react with a further malonyl-CoA unit unless the NADPH-dependent reduction reaction has occurred. This is consistent with a proposed *processive* mechanism of polyketide formation, in which processing of the keto groups occurs during

Figure 1.3 Hypothetical reaction pathways for 6-methylsalicylic acid biosynthesis in *Penicillium patulum*.

The upper pathway from acetyl- and malonyl-CoA to 6-methylsalicylic acid (6MSAS) is adapted from Bedford *et al.* (1995) and follows the route proposed by Dimroth *et al.* (1970), in which triacetic acid (TAA) reduction is followed by dehydration.

The lower pathway from TAA to 6MSAS is an alternative proposed by Schorr *et al.* (1994) wherein dehydration occurs at the tetracetic rather than the triacetic acid level.

E, PKS enzyme; S_K, ketosynthase active cysteine; S_{ACP}, ACP prosthetic group; KS, ketosynthesis; AT, acetyl transfer; MT, malonyl transfer; *, transacylation from ACP to ketosynthase active site; KR, ketoreduction; CYC, cyclisation.


acetyl-CoA
AT
malonyl-CoA
MT
KS
CO2
malonyl-CoA
MT
*
KS
CO2
TAA
KR
NADPH
DH
H20
*
MT
malonyl-CoA
KS
CO2
CYCLIZATION
H20
RELEASE
H20
6-MSA
KR
NADPH
MT+KS
DH+CYC
TAA
6-MSA

the assembly of the polyketide chain (O'Hagan, 1991; Spencer & Jordan, 1992a). However, the subsequent dehydration of the keto group seems to occur after condensation of the third malonyl-CoA onto the carbon chain, along with cyclisation of the tetraketide intermediate (figure 3 and Schorr *et al.*, 1994). Release of the completed 6MSA molecule from the enzyme is not inactivated by treatment with the serine protease inhibitor PMSF and is therefore unlikely to occur *via* the action of a serine-dependent thioesterase, as in the structurally similar mammalian FAS (Stoops & Wakil, 1981). Product release from 6MSAS may involve a different class of thioesterase or another mechanism altogether (Jordan & Spencer, 1993; Schorr *et al.* 1994; Spencer & Jordan, 1992a).

In common with the fatty acid synthases, 6MSAS incorporates malonyl-CoA into the reactive polyketide intermediates by a highly stereospecific process. The NADPH-dependent ketoreductase has been shown to be *Si*-face specific, which is the same stereoselectivity exhibited by FAS (Schorr *et al.*, 1994). The steric course of the reaction has been investigated using isotopically-labelled chiral malonyl Co-A derivatives which were incorporated into 6-methylsalicylic acid using purified 6MSAS (Spencer & Jordan 1992b). An understanding of the stereochemical events involved will be necessary to elucidate the mechanistic course of the reaction at each stage of the synthesis (Jordan & Spencer, 1993; Spencer & Jordan, 1992b).

1.3.3.2.2 Orsellinic acid synthase

The only other fungal PKS for which purification, characterisation and mechanistic studies have been reported is orsellinic acid synthase, from *Penicillium cyclopium* (Jordan & Spencer, 1993; Spencer & Jordan, 1992). Orsellinic acid, a tetraketide, is structurally very similar to 6MSA and is synthesised from the same substrates, but it is a non-reduced polyketide. As a consequence orsellinic acid synthase lacks the reductase and dehydratase activities found in 6MSAS and, consistent with its smaller size (130 kDa), the protein appears to lack the corresponding functional domains. In contrast to 6MSAS, orsellinic acid synthase does not produce the by-product triacetic acid lactone, suggesting that the missing reductase and/or dehydratase catalytic sites may be involved in release of the triketide (Jordan & Spencer, 1993). As with 6MSAS, orsellinic acid synthase appears to exert a high degree of steric control over the reactions it catalyses.

1.3.3.3 Higher plant polyketide synthase enzymes

Two closely related higher plant PKS enzymes have been purified from cell-suspension cultures and characterised. Stilbene synthases (STS) are involved in the biosynthesis of stilbene-type phytoalexins and different types of STS can be distinguished by their substrate specificities (Schröder *et al.*, 1988). Resveratrol synthase (RS) from *Arachis hypogaea* (peanut) is an STS which catalyses the synthesis of the phytoalexin resveratrol from an aromatic 4-coumaroyl-CoA starter unit and three molecules of malonyl-CoA (Schöppner & Kindl, 1984). Chalcone synthases (CHS) are ubiquitous enzymes in higher plants that perform a key reaction in flavonoid biosynthesis, utilising the same substrates as RS (Hopwood & Sherman, 1990; Schröder *et al.*, 1996). A CHS which produces the flavonone naringenin has been purified from *Petroselinum hortense* (parsley) (Kreuzaler & Hahlbrock, 1975). RS was found to have a specific requirement for malonyl-CoA and, while it would accept various CoA thioesters of aromatic carboxylic acids in place of 4-coumaroyl-CoA, these gave greatly reduced yields of stilbene product. CHS also exhibits a narrow specificity as far as aromatic CoA esters are concerned but will utilise aliphatic starter molecules at a similar efficiency to that observed for 4-coumaroyl-CoA (Schüz *et al.*, 1983). Taking into account the similarities between CHS and the ß-ketoacyl-ACP synthase of FAS, this may indicate that the CHSs have evolved from an enzyme acting on aliphatic substrates (Schüz *et al.*, 1983). Although STS and CHS enzymes catalyse the same sequence of condensations, differential folding of the carbon chain prior to cyclisation will lead either to an aldol reaction and a stilbene product or to a claisen ester condensation and a chalcone product (Schöppner &Kindl, 1984). Both enzymes have a similar subunit molecular weight and exhibit a dimeric active form. These enzymes lack both ACP and acyltransferase activities; they utilise acyl CoA thioesters directly as substrate rather than ACP-linked units and are essentially equivalent to the ß-ketoacyl synthase III of the *Escherichia coli* FAS. Recent characterisations of related enzymes suggest that CHS and STS are members of an enzyme superfamily which makes use of widely varying starter units (Junghanns *et al.*, 1995; Schröder *et al.*, 1996).

The partially reduced polyketide 6-hydroxymellein is a direct precursor of an antimicrobial agent produced by carrot cells as a defensive response to pathogen attack. 6-Hydroxymellein synthase, a polypeptide with several functional domains, has been partially (240-fold) purified and characterised (Kurosaki *et al.*, 1993). The NADPH-dependent

ketoreduction involved in 6-hydroxymellein biosynthesis exhibits identical stereospecificities to those of the FAS ß-ketoacyl reductase.

1.3.4 Molecular programming in polyketide synthases

The great diversity in polyketide structure is due to the variations on the mechanistic theme of fatty acid biosynthesis that may be employed by the PKS enzyme. These may include varying degrees of reduction, use of chain extension units other than malonyl-CoA, methylation and cyclisation of the carbon chain and chain termination. To ensure consistent synthesis of the correct polyketide, these activities must be deployed at the correct stage in the biosynthetic pathway; this regulation is known as molecular programming.

In the case of the modular PKS enzymes, a separate functional domain is responsible for catalysing each stage of the synthesis. Molecular programming and hence the structure of the final product is mediated by the specificity of each synthase unit for the correct intermediate and extender unit (Donadio *et al.*, 1991). In the case of the type I and type II PKS, only a single set of functional domains is present as in the FAS. To introduce complexity into the polyketide product the PKS must utilise these active sites selectively in a defined order. An example of this can be seen in the case of 6-methylsalicylic acid biosynthesis where NAPDH-dependent reduction of the intermediate occurs only once, at the C6 stage, during the three cycles of chain extension catalysed by 6MSAS.

How this regulation is accomplished is not fully understood as yet, but some useful inferences have been drawn from the comparison of 6MSAS with orsellinic acid synthase which produces a structurally similar but non-reduced tetraketide (Jordan & Spencer, 1993). Although the enzymes share a C6 intermediate, under conditions of NADPH starvation, where no reduction will occur, orsellinic acid will not be produced by 6MSAS (as discussed above). This suggests that either 6MSAS will not incorporate a further malonyl-CoA unit into the unreduced C6 intermediate or it is unable to cyclise an unreduced C8 carbon chain. Hence 6MSAS is programmed to produce only 6-methylsalicylic acid as an end product (Jordan & Spencer, 1993). 6MSAS and the fatty acid synthases will release their common C6 intermediate as the triacetic acid lactone by-product. This background reaction does not occur in the case of orsellinic acid synthase, leading to speculation that the ketoreductase or dehydratase domains missing from orsellinic acid synthase may be involved in the triacetic acid release mechanism (Jordan & Spencer, 1993).

The FAS ketosynthase enzymes of *E. coli* and spinach leaf have been shown to exhibit greatly reduced activity on substrates with a chain length over a certain size (reviewed by Hopwood & Sherman, 1990). This suggests an essential role for these proteins in chain length determination by some FAS and PKS systems. At most stages in polyketide biosynthesis the active sites of the PKS interact with an ACP-bound intermediate; this small protein domain could have a major influence on the programming of the PKS. Enzyme attachment *via* the ACP could effect channelling of intermediates along the correct pathway by imposing conformational constraints on the stereochemistry of the catalytic activities (Hopwood & Sherman, 1990; Hutchinson & Fujii, 1995; Jordan & Spencer, 1993).

1.4 POLYKETIDE SYNTHASE GENES

1.4.1 Bacterial polyketide synthase genes

Actinomycete polyketide synthase genes involved in the biosynthesis of aromatic polyketides, such as the *Streptomyces coelicolor* A3(2) *act* genes for actinorhodin biosynthesis, show a type II architecture (figure 4), with a cluster of open reading frames encoding the various iteratively-used components of the PKS (Fernández-Moreno *et al.*, 1992; Hallam *et al.*, 1988; Hopwood & Khosla, 1992; Hopwood & Sherman, 1990; Hutchinson & Fujii, 1995). All type II gene clusters identified so far include a set of three genes encoding the 'minimal PKS': the ketosynthase, which also carries a putative acyltransferase domain (Fernández-Moreno *et al.*, 1992; Shen & Hutchinson, 1993b); the chain length determining factor (CLF) and an ACP (McDaniel *et al.*, 1995). The CLF is situated immediately adjacent to the ketosynthase in each case and is a highly homologous protein lacking the KS active site cysteine residue (McDaniel *et al.* 1993). Ketoreductase, methyltransferase and a cyclase/aromatase enzyme may also be present (Hutchinson & Fujii, 1995; McDaniel *et al.*, 1995). The early discovery that some of these genes would extensively cross-hybridise (Malpartida *et al.*, 1987), together with the close linkage of these genes, has facilitated the isolation of at least 13 clusters from various *Streptomyces* and *Saccharopolyspora* species (Hutchinson & Fujii, 1995). A strong similarity is apparent between these clusters in the linear arrangement and direction of transcription of the enzyme genes.

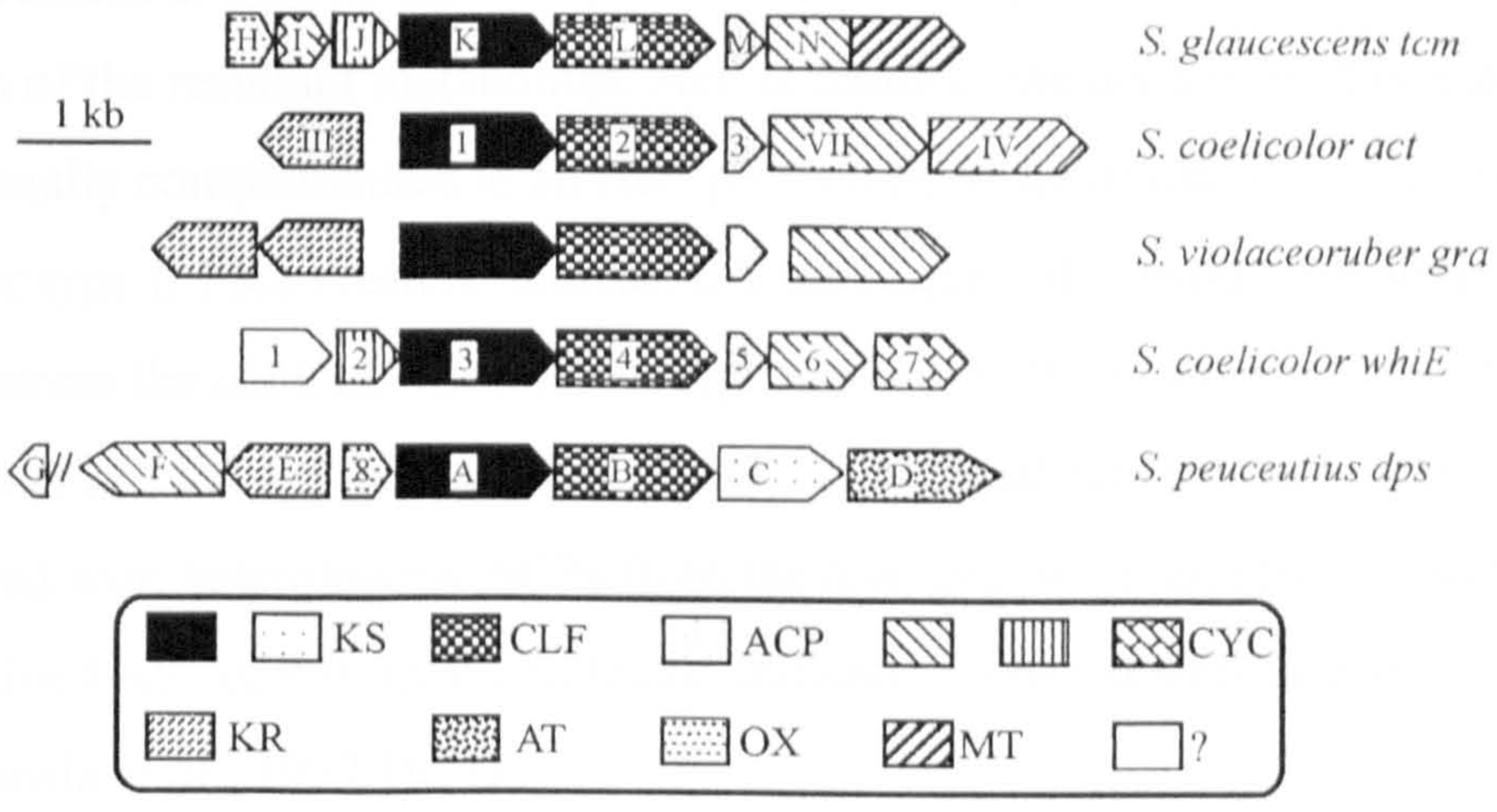


Figure 1.4 Type II PKS gene clusters.

Genes are shaded to indicate known functions: KS, ketosynthase; CLF, chain length determining factor; ACP, acyl carrier protein; KR, ketoreductase; AT, acyltransferase; CYC, cyclase; OX, oxidation; MT, methyltransferase; ?, function unknown.

Slashes indicate that open reading frames intervene between the genes shown.

Adapted from Hutchinson & Fujii (1995).

Some type II acyltransferase genes seem to be 'shared' between FAS and PKS complexes, including the *fabD* homologue involved in actinorhodin biosynthesis in *Streptomyces coelicolor* (Revill *et al.*, 1995) and the *fabD* gene from *S. glaucescens* which may be involved in tetracenomycin biosynthesis (Summers *et al.*, 1995).

A number of published papers describe experiments designed to determine the specificities of the individual enzymes making up the type II PKS, *via* the complementation of mutants or the replacement of PKS components with their counterparts from other type II PKS and examination of the resultant metabolites. Act^- mutants in the *act* ketosynthase gene have been functionally complemented to an Act^+ phenotype by ketosynthase genes from the *gra*, *whiE* and *otc* type II PKS clusters, indicating a fairly general substrate requirement for this enzyme, whereas the *act* CLF was only complemented by its counterpart from the *gra* cluster (Kim *et al.*, 1994; Sherman *et al.*, 1992). Functional replacement of the *act* ACP has been achieved with heterologous ACPs from the *tcm*, *gra*, *whiE* and *fren* clusters and even with a putative FAS ACP from *S. erythraea*, indicating a broad substrate specificity for this enzyme (Khosla *et al.*, 1992,1993).

A type II PKS gene involved in the biosynthesis of the polyketide component of the phytotoxin coronatine has recently been identified in *Pseudomonas syringae* (Penfold *et al.*, 1996) and two unusual features of this PKS were noted. It is involved in the synthesis of a complex polyketide and is the first type II PKS to have been identified with such a role, in addition the putative dehydratase is related to the FAS dehydratase domain rather than the putative actinomycete type II PKS dehydratase.

The actinomycete macrolide antibiotics are produced by modular/type I PKS enzymes. These are giant multifunctional proteins exemplified by the 6-deoxyerythronolide B synthase complex (DEBS), which is involved in the biosynthesis of erythromycin A by *Saccharopolyspora erythraea* (Bevitt *et al.*, 1992; Donadio *et al.*, 1991). The genes encoding DEBS are located on a 30-35 kb segment of chromosome designated *eryA*. They consist of three large ORFs, each containing two modules which encode the unique catalytic sites required for two of the six elongation steps in the synthesis of the full-length carbon chain (figure 5). Domain inactivation studies have confirmed that the 5'-3' order of the six modules is colinear with the steps in assembly of the 6-deoxyerythronolide B (6dEB) backbone; this has led to the proposal of a non-iterative model proposed for DEBS, wherein each module is involved in a single elongation step (Donadio *et al.*, 1993,1991).

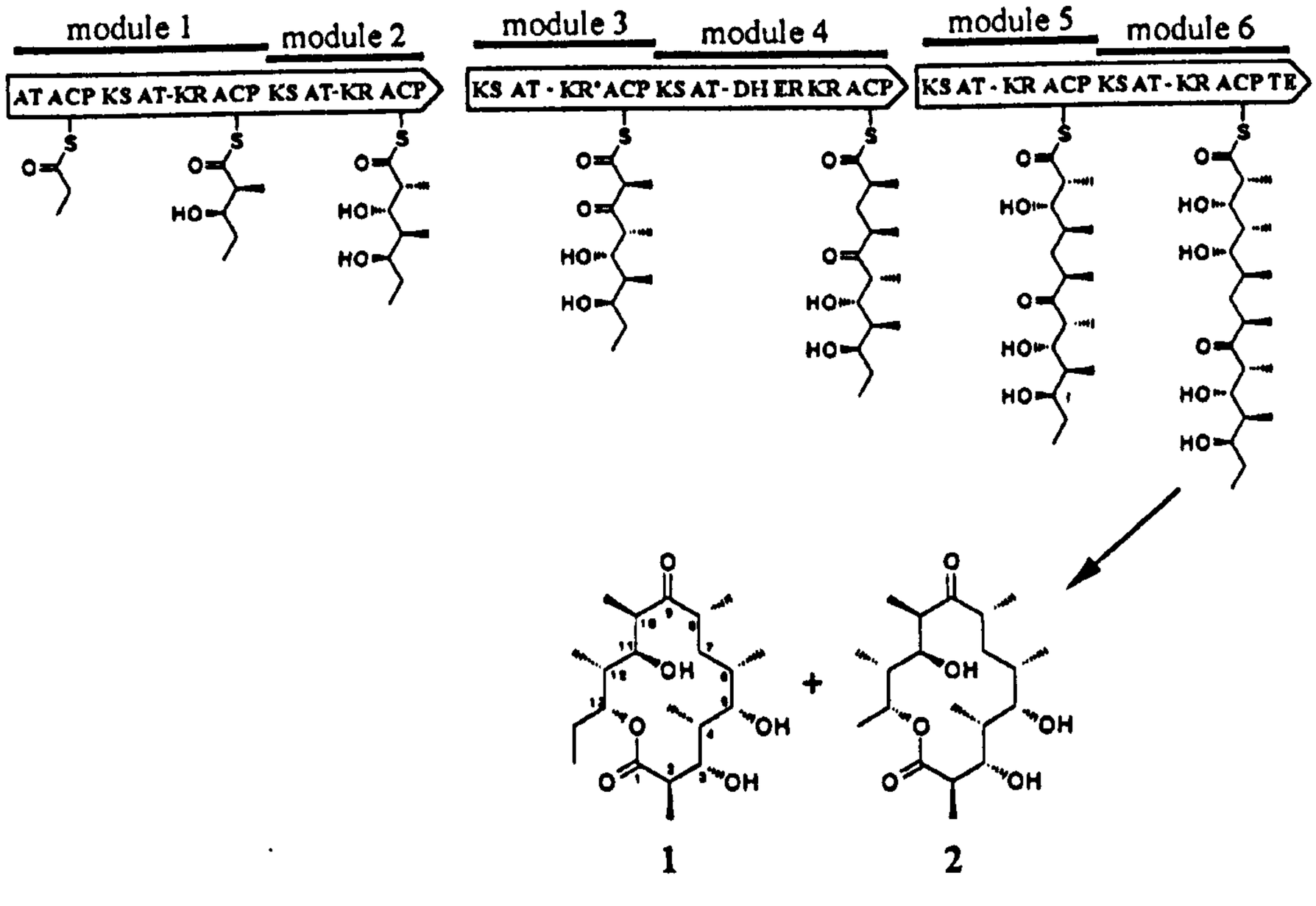


Figure 1.5 Genetic model for 6-deoxyerythronolide B synthase (DEBS).

Each module posesses a ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) domain, in addition to a reductive segment consisting of an interdomain linker (-), a ketoreductase (KR) and, in module 4, dehydratase (DH) and enoylreductase (ER) domains. The KR domain in module 3 (KR^*) lacks reductase activity. Lactone formation is catalysed by the thioesterase domain of module 6.

When expressed heterologously in *Streptomyces coelicolor*, DEBS produces 6-deoxyerythronolide B (**1**) as well as 8,8a-deoxyoleandolide (**2**) (Kao *et al.*, 1994).

Adapted from Bedford *et al.* (1996).

A similar modular structure has been identified by DNA sequence analysis in other actinomycete type I PKS involved in the assembly of other complex polyketides, including the macrolide antiparasitic avermectins (MacNeil *et al.*, 1994) and the macrocyclic immunosupressant rapamycin (Schwecke *et al.*, 1995). The three clustered PKS genes (*rapA/B/C*) required for rapamycin biosynthesis in *Streptomyces hygroscopicus* together encode a total of 14 modules on three giant polypeptides (RAPS1/2/3), making this the most complex multienzyme system identified so far (Aparicio *et al.*, 1996; Schwecke *et al.*, 1995). In contrast to DEBS, many catalytic domains not required for the action of a particular RAPS module are present in an inactive form, rather than being deleted (Aparicio *et al.*, 1996). Genes encoding putative modular type I PKS have been identified in *Bacillus subtilis* (Albertini *et al.*, 1995; Scotti *et al.*, 1993) and also in *Mycobacterium leprae* (Hutchinson & Fujii, 1995; Robison & Smith, 1994), although the products are unknown in both cases. Polyether antibiotics such as tetronasin are derived from acetate, propionate and butyrate units, an origin so far associated with the type I PKS, although no such PKS has as yet been identified (Linton *et al.*, 1994).

1.4.2 Fungal polyketide synthase genes

Fungal PKS may be considered as a hybrid category between the two classes of bacterial PKS; a single gene encodes all the active sites required for polyketide biosynthesis, which are used iteratively. Intron structure and functional domains of cloned fungal PKS genes are shown in figure 1.6. The *Penicillium patulum* 6MSAS gene was cloned by immunological screening of a genomic DNA expression library in 1990 (Beck *et al.*, 1990). A 7.7 kb genomic DNA clone containing the 3' terminus of this gene was subsequently isolated using a degenerate oligonucleotide probe (Wang *et al.*, 1991). Within the predicted protein product of 1774 amino acids, ketosynthase, ketoreductase, acyltransferase and ACP domains were identified by homology to animal FAS (Beck *et al.*, 1990; Wang *et al.*, 1991) and a dehydratase domain by homology to the *E. coli* FabA dehydratase (Bevitt *et al.*, 1992). Both amino acid sequence conservation in active domains and overall structural similarities are much higher when 6MSAS of *P. patulum* is compared to the type I animal FAS than when it is compared to its own homologous α/ß subunit type I FAS, suggesting that the PKS and FAS in this organism have arisen by convergent rather than divergent evolution (Beck *et al.*, 1990; Hopwood & Sherman, 1990; Wang *et al.*, 1991). There is no evidence of either an enoylreductase nor a thioesterase domain. The absence of an

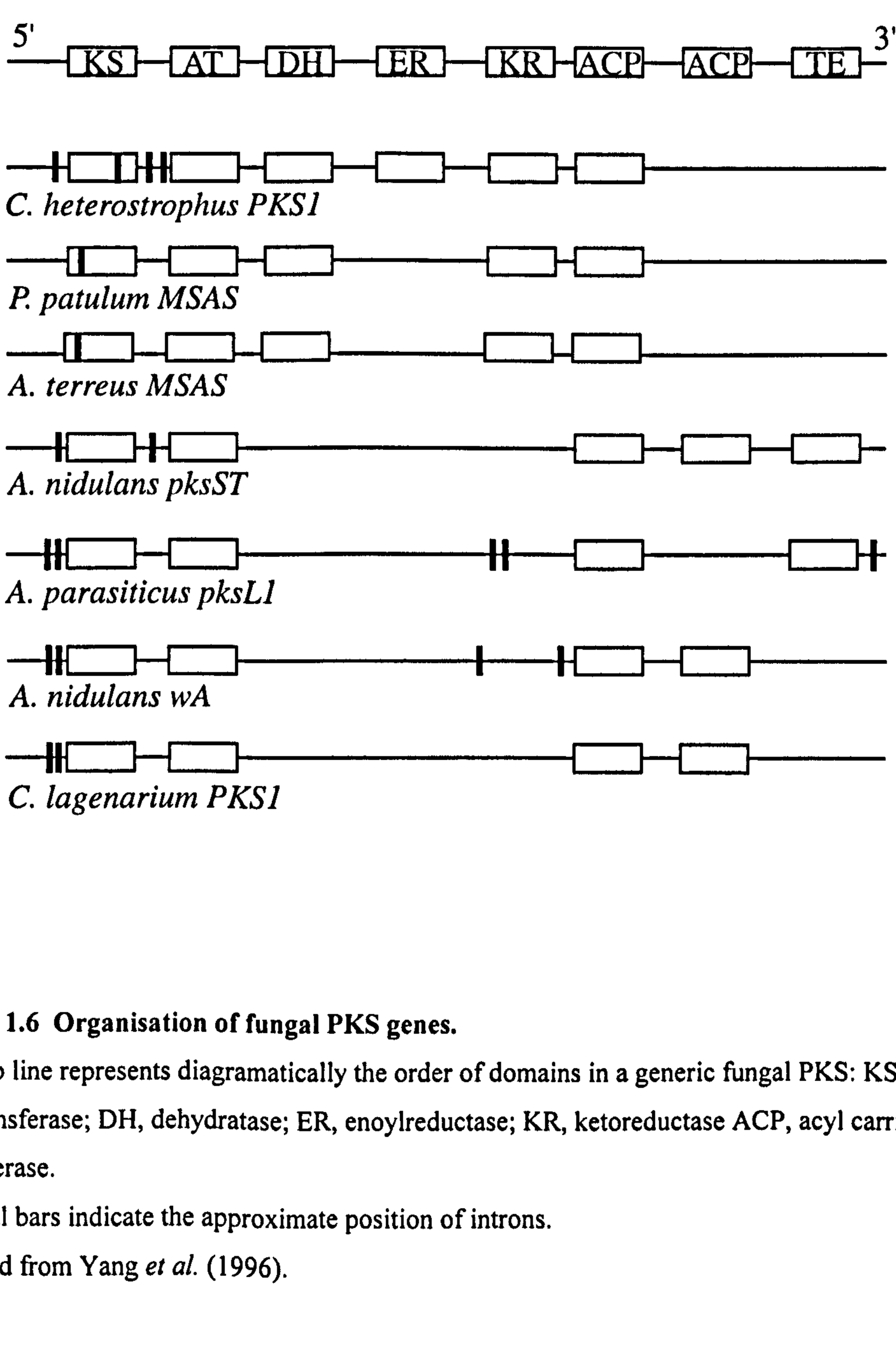


Figure 1.6 Organisation of fungal PKS genes.

The top line represents diagramatically the order of domains in a generic fungal PKS: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase ACP, acyl carrier protein; TE, thioesterase.

Vertical bars indicate the approximate position of introns.

Adapted from Yang *et al.* (1996).

enoylreductase is consistent with the biosynthetic scheme proposed for 6MSA (figure 3), and the lack of any recognisable thioesterase domain is consistent with results from studies of the MSAS enzyme (Jordan & Spencer, 1993; Schorr *et al.*, 1994; Spencer & Jordan, 1992a). A homologous gene from *Aspergillus terreus* has been cloned using the ketosynthase domain of *Penicillium patulum MSAS* as a probe and shown to encode MSAS by heterologous expression in *A. nidulans* (Fujii *et al.*, 1996; Pazoutova *et al.*, 1995). The structure of this gene is very similar to that of *P. patulum MSAS*; a small 5' exon is separated from the body of the coding sequence by a short intron (Fujii *et al.* 1996). A reconstructed version of the *P. patulum MSAS* gene has been expressed in *Streptomyces coelicolor* strain CH999, resulting in the production of significant amounts of 6-methylsalicylic acid (Bedford *et al.*, 1995). The use of this genetically well characterised system for expression should allow the use of genetic manipulations such as site-directed mutagenesis for the elucidation of structure-function relationships in MSAS and other PKS (Bedford *et al.*, 1995).

The *A. nidulans wA* gene, required for synthesis of a green pigment found in the mature conidial wall, was cloned by complementation of spore pigment mutants and characterised as a PKS by DNA sequence analysis (Mayorga & Timberlake, 1990,1992; Tilburn *et al.*, 1990). A polypetide of 1986 amino acids is encoded by *wA,* within which ketosynthase, acyltransferase and two ACP motifs have been identified by homology to other PKS and animal FAS (Mayorga & Timberlake, 1992).

A PKS gene required for aflatoxin biosynthesis in *A. parasiticus* (*pksA/pksL1*) has been identified and cloned independently by three groups, using targeted gene disruption and homologous PCR-derived probes (Chang *et al.*, 1995; Feng *et al.*, 1995; Trail *et al.*, 1995). Within the polypeptide of 2109 amino acids encoded by *pksA/pksL1* four functional domains have been identified: ketosynthase, acyltransferase, thioesterase and ACP (Chang *et al.*, 1995; Feng *et al.*, 1995). None of the domains involved in reduction of keto groups (ketoreductase, dehydratase, enoylreductase) can be identified and this is consistent with the proposed use of a hexanoate starter for chain building (Brobst & Townsend, 1994; Chang *et al.*, 1995), provided by a specialised α/ß subunit type FAS (Mahanti *et al.* 1996; Watanabe *et al.*, 1996). The pksA/pksL1 predicted amino acid sequence shows a high similarity to *A. nidulans wA* and the first two introns in the two genes interrupt the coding sequence at the same sites, pksA/pksL1 however has only a single ACP domain compared

to the two identified in wA and also has about 130 extra amino acids at the C-terminus (Chang *et al.*, 1995; Feng *et al.*, 1995).

In addition to *wA*, a second PKS gene has been isolated from *A. nidulans*; *pksST* is required for biosynthesis of sterigmatocystin (ST), a product related to aflatoxin, and was cloned via the use of deletion mutagenesis to identify genes associated with ST biosynthesis (Yu & Leonard, 1995). The predicted gene product, 2181 amino acids in length, is very similar in sequence (77%) to *pksA/pksL1* of *A. parasiticus*, which presumably catalyses the formation of the same intermediate (Feng & Leonard, 1995; Yu & Leonard, 1995). The two genes also possess the same functional domains, although *pksST* has a duplicated ACP domain.

Biosynthesis of the melanin required for infection by many plant pathogens is via the 1,8-dihydroxynaphthalene pathway and the first stage in this process is polyketide synthesis (Bell & Wheeler, 1986). *PKS1,* the PKS gene required for melanin biosynthesis in *Colletotrichum lagenarium,* has been cloned by complementation of an albino mutant (Takano *et al.*, 1995). The 2187-residue amino acid sequence encoded by *PKS1* is extremely homologous to the predicted product of *A. nidulans wA* throughout its length and also exhibits the duplicated ACP domain; homology to *P. patulum* 6MSAS is restricted to active domains (Takano *et al.*, 1995).

Cochliobolus heterostrophus Race T uniquely produces a family of linear long-chain polyketides, collectively known as T-toxin, which are responsible for its virulent pathogenicity towards Texas male-sterile maize (Siedow *et al.*, 1995). A PKS gene (*PKS1*) has been identified at the *Tox1* locus responsible for T-toxin production and disruption of *PKS1* has been shown to eliminate both T-toxin production and fungal virulence (Yang *et al.*, 1996). Ketosynthase, acyltransferase, dehydratase, enoylreductase, ketoreductase and ACP domains can be identified on the 2530-amino acid polypeptide encoded by *PKS1,* which shows a higher similarity to various bacterial PKS and animal FAS than to other fungal PKS (Yang *et al.*, 1996). Although the G+C content and presence of introns in *PKS1* are consistent with a fungal origin, the codon usage within the ORF of *PKS1* differs from the majority of fungal genes, including most *Cochliobolus* genes (Yang *et al.*, 1996). Together with the resemblance of PKS1 to bacterial PKS, this suggests that this gene may have been acquired by horizontal transfer from a prokaryotic source; such an origin has been proposed previously for genes from *Cochliobolus* and other fungal species involved in pathogenicity (Masel *et al.*, 1993; Panaccione *et al.*, 1992; Yang *et al.*, 1996).

A putative PKS gene, cloned from *Aspergillus terreus*, is known to be involved in biosynthesis of lovastatin, a highly reduced, methylated polyketide (Davis *et al.*, 1994). Sequence analysis indicated that the gene possesses the expected full set of FAS reducing activities and an integral methyltransferase domain.

1.4.3 Higher plant polyketide synthase genes

Chalcone synthase genes have been cloned from a number of higher plant species and found to be a highly conserved in amino acid sequence along most of the length of the protein (Niesbach-Klösgen *et al.,* 1987). All CHS genes sequenced so far possess an intron at exactly the same point and this intron is conserved in genes for resveratrol synthase (Schröder *et al.*, 1988). Deduced amino acid sequences for CHS genes from different plant species are generally greater than 80% similar and the similarity between CHS and RS amino acid sequences is generally greater than 70% (Schröder *et al.*, 1988), indicating that the two types of enzyme are closely related. Neither enzyme resembles the FAS or PKS ketosynthases, which suggests that these two types of condensing enzyme are evolutionarily unrelated (Hopwood & Sherman, 1990). In both CHS and RS the active cysteine thiol is not part of a motif resembling the active site of other FAS or PKS KS enzymes, confirming the notion of a separate development for the CHS and RS enzymes (Lanz *et al.*, 1991). Acridone synthase (ACS), an enzyme closely related to both CHS and RS which employs N-methylanthraniloyl-CoA as a starter unit, has been cloned from *Ruta graveolens* and its deduced amino acid sequence shows greater than 60% similarity to CHS enzymes from other species; only a few subtle changes in the primary sequence of CHS may be required to change the substrate specificity to that of ACS (Junghanns *et al.*, 1995).

1.4.4 The use of recombinant polyketide synthase enzymes as a source of novel polyketides.

As a consequence of the pharmacological value of many polyketide natural products, a considerable amount of research is being directed towards the possibility of generating novel compounds via genetic engineering of biosynthetic pathways. Early experiments involved the transfer of partial or complete biosynthetic gene clusters between different polyketide-producing strains of *Streptomyces* and yielded the novel polyketides dihydrogranatirhodin and the mederrhodins (Hopwood *et al.*, 1985). This empirical approach involved "tailoring enzymes" rather than the PKS-catalysed steps, as did the more

targeted genetic manipulations of *Streptomyces ambofaciens*, wherein the *carE* acylase gene was introduced from *S. thermotolerans* to acylate the native macrolide product spiramycin (Epp *et al.*, 1989; Hopwood *et al.*,1985; McDaniel *et al.*, 1995). Subsequent studies aimed at the production of novel polyketides have focused on the PKS enzymes and have generally involved either rational design of recombinant modular type I PKS or combinatorial biosynthesis using type II systems. Both of these strategies are discussed below.

The rational approach to producing novel macrolide compounds is based on the idea that each enzymic domain within a modular PKS should affect only a single step in biosynthesis (reviewed by Donadio *et al.*, 1991; Katz & Donadio, 1993). Elucidation of the correlation between polyketide structure and PKS gene organisation should facilitate selective genetic engineering of the PKS, resulting in a predictable polyketide product. This approach has been used to generate erythromycin analogues via the deletion of active sites within the DEBS PKS complex, each deletion removing a step in a specific reductive cycle (Bedford *et al.*, 1996; Donadio *et al.*, 1991,1992). Several research groups have repositioned the region of the *eryAIII* gene encoding the chain-terminating thioesterase domain of DEBS3 to the carboxyl-terminus of *eryAI,* to produce a chimaeric gene encoding a product known as "DEBS1+TE" (figure 7). This recombinant PKS has been expressed heterologously in *Streptomyces coelicolor* strain CH999 (Kao *et al.*, 1995; Pieper *et al.*, 1995a,b) and also exchanged for the chromosomal copy of *eryA* in *Saccharopolyspora erythraea* by homologous integration (Cortes *et al.*, 1995). Both approaches resulted in a good yield of the predicted triketide lactone, confirming the cyclase activity of the TE domain and proving that the multienzyme components of modular PKS are capable of functioning independently. Successful incorporation of butyryl-CoA into a triketide lactone by cell-free DEBS1+TE suggested that similar *in vitro* systems may provide a means for the controlled synthesis of polyketides which would be unobtainable by *in vivo* methods (Pieper *et al.*, 1995b).

The clear correlation between active sites and product structure and the relaxed specificity for starter and, to a lesser degree, extender units indicate that modular bacterial PKS offer greater potential for the generation of novel compounds by genetic engineering than do the type II PKS (Tsoi & Khosla, 1995). In contrast, current knowledge of the combinatorial potential in modular PKS is relatively limited when compared to that concerning the type II complexes (Khosla & Zawada, 1996). Combinatorial biosynthesis involves manipulation of

PKS genes so that they may be recombined with genes from other such systems, leading to the production of novel polyketides (reviewed by Khosla & Zawada, 1996; McDaniel *et al.*, 1995). A specially designed host-vector system comprising *Streptomyces coelicolor* strain CH999, from which the *act* PKS genes have been deleted, and the shuttle plasmid pRM5 can be used to construct and express complete recombinant minimal PKS gene clusters (McDaniel *et al.*, 1993). This, in conjunction with emerging information on the function and specificities of the various type II enzyme subunits, has enabled the rational design and synthesis of novel aromatic polyketides by recombinant assembly of PKS component enzymes (McDaniel *et al.*, 1995).

In addition to the actinomycetes, the fungi are potentially a valuable source of genes for use in the rational design and generation of combinatorial libraries of recombinant PKS. Fungal polyketides show several unusual features which indicate a diversity in the range of fungal PKS not found in the bacterial enzymes currently exploited (Bedford *et al.*, 1995). These features include the wide range of chain lengths found in aromatic polyketides (Simpson, 1985), the *S*-adenosyl-L-methionine origin of methyl groups, unusual chain cyclisation in compounds such as bikaverin and the production of wholly or largely reduced polyketides such as mevinolin (O'Hagan, 1991). Also, as single multifunctional polypeptides, fungal PKS are particularly suitable for heterologous gene expression (Bedford *et al.*, 1995).

1.5 ENZYMIC MODIFICATION OF THE POLYKETIDE SYNTHASE PRODUCT.

The polyketide backbone synthesised by a PKS is normally processed further by a series of enzyme-mediated reactions, such as oxidation, reduction or alkyklation, in a biosynthetic pathway. These 'tailoring' reactions will often impart to the polyketide its characteristic biological activity (Hutchinson & Fujii, 1995). In general, enzymes of secondary metabolism have proved very difficult to purify in an active form, and this has retarded study of these tailoring enzymes (Bennett, 1995; Dutton, 1988; Robinson, 1991). In several cases, however, the genes encoding these enzymes have been cloned and the predicted amino acid sequences have provided much information. Three important classes of post-PKS modifications to the carbon chain, and the enzymes effecting them, are discussed below (reviewed by Hutchinson & Fujii, 1995).

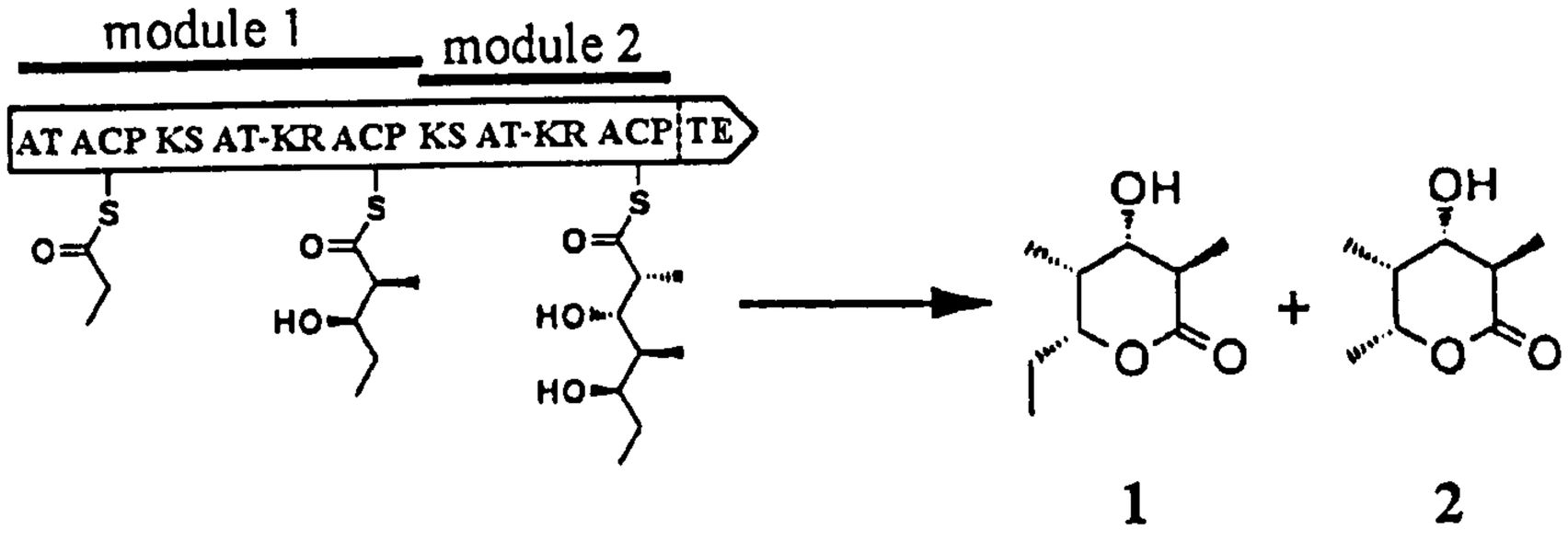


Figure 1.7 Genetic model for DEBS1+TE.

This recombinant bimodular PKS, consisting of DEBS modules 1 and 2 attached to the thioesterase (TE) domain from module 6, produces the triketide lactones **1** and **2** when expressed in *Streptomyces coelicolor* (Kao *et al.*, 1995; Pieper *et al.*, 1995a,b).

1.5.1 Reduction

Several bacterial PKS complexes include ketoreductases (Hutchinson & Fujii, 1995). The *act*III gene from the actinorhodin biosynthetic cluster encodes a ketoreductase homologous to known oxidoreductases (Hallam *et al.*, 1988). Biosynthetic experiments involving expression of recombinant PKS genes have shown that the ActIII ketoreductase will reduce carbon chains of varying lengths, always at the C9 carbonyl, and probably acts after completion of the complete chain ,i.e. non-processively (Fu *et al.*, 1993).

The aflatoxin biosynthetic pathway includes several reduction steps (Dutton, 1988; Hutchinson & Fujii, 1995). Two ketoreductase genes in the aflatoxin cluster and their homologues in the sterigmatocystin pathway have been cloned and characterised in *Aspergillus parasiticus*; *A. flavus* and *A. nidulans*: *ver-1/verA/stcU* act on versicolorin A and *norA/nor-1/stcE* act on norsolorinic acid (Brown *et al.*, 1996; Cary *et al.*, 1996; Keller & Adams, 1995; Yu *et al.*, 1995). The predicted proteins show significant conservation of size and amino acid sequence with higher plant and bacterial ketoreductases (e.g. ActIII) involved with the modification of ring structures (Skory *et al.*, 1992; Trail *et al.*, 1994). Two distinct reductase enzymes acting on versiconal hemiacetal acetate have been purified to homogeneity from *A. parasiticus* but a gene is yet to be isolated (Matsushima *et al.*, 1994).

Fungal DHN melanins, involved in protection of propagules against environmental stress and phytopathogenicity, are synthesised from 1,8-dihydroxynaphthalene (DHN), which is in turn synthesised from 1,3,6,8-tetrahydroxynaphthalene, the direct product of a PKS (Bell & Wheeler, 1986). This pentaketide is subsequently reduced to scytalone by a NADPH-dependent reductase, which has been purified from *Magnaporthe grisea* (Vidal-Cros *et al.*, 1994). Antibodies raised against the reductase were used to isolate a full-length cDNA which, when expressed as a ß-galactosidase fusion protein, was found to catalyse a second reduction in the DHN pathway, forming vermelone from 1,3,8-trihydroxynaphthalene (Vidal-Cros *et al.*, 1994). The deduced amino acid sequence was highly conserved with that of *A. parasiticus ver-1* and showed similarities to numerous other oxidoreductases in the short-chain dehydrogenase family. Emodin deoxygenase, also involved in reduction of an aromatic ring, has been purified from *Pyrenochaeta terrestris* and found to consist of two proteins acting synergistically (Anderson *et al.*, 1990; Ichinose *et al.*, 1993).

1.5.2 Oxidation

Common oxidative modifications of polyketides include hydroxylation of aromatic or reduced polyketides, conversion of hydroquinols to quinones, oxidative cleavage of double bonds or aromatic rings and oxidative coupling reactions. One of the most important types of enzyme involved in secondary metabolism are the cytochrome P450 monooxygenases, which incorporate a single atom of molecular oxygen into their substrate (the other atom being reduced by NADPH). These enzymes are often involved in the hydroxylation of polyketides and several cytochrome P450 genes have been found in secondary metabolic gene clusters.

Genes for two *Saccharopolyspora erythraea* P450 enzymes involved in erythromycin A biosynthesis have been identified within the erythromycin biosynthetic gene cluster (Stassi *et al.*, 1993; Weber *et al.*, 1991). *EryF,* identified by a targeted disruption of the *S. erythraea* chromosome, encodes 6-deoxyerythronolide B hydroxylase, which acts on the macrolactone product of the PKS encoded by the *EryA* locus (Weber *et al.*, 1991). The enzyme has been purified from *S. erythraea* and over-expressed in recombinant *E. coli*, allowing extensive modelling of the substrate-enzyme complex (Andersen *et al.*, 1993; Shafiee & Hutchinson, 1987). As with most cytochrome P450 systems, 6-deoxyerythronolide B hydroxylase requires two non-specific electron transport proteins for catalytic activity: ferredoxin and NAD(P)H:ferredoxin oxidoreductase, encoded by the genes *fdxA* and *forA*. The normal arrangement for actinomycete P450 systems involved in primary metabolism is for the genes encoding the electron transport proteins to be clustered with the P450 genes (O'Keefe & Harder, 1991). *fdxA* and *forA* are not closely linked to the erythromycin biosynthetic gene cluster and hence are not linked to e*ryF* (Donadio and Hutchinson 1991; Zotchev & Hutchinson, 1995). This non-linkage appears to be typical of P450 systems dedicated to secondary metabolism, possibly indicating that expression of electron transport proteins is not co-ordinated with that of the P450 (Zotchev & Hutchinson, 1995). One of the final reactions in erythromycin A biosynthesis is a C-12 hydroxylation of erythromycin D, catalysed by the cytochrome P450 EryK. The *eryK* gene, identified by comparative analysis and gene disruption, has been cloned from *S. erythraea* (Stassi *et al.*, 1993). Over-expression in *E. coli* produced insoluble inclusion bodies which were solublised in 8M urea and reconstituted to an active holo-enzyme, allowing characterisation of the substrate specificity of the enzyme (Lambalot & Cane, 1995).

Four P450 monooxygenase genes (*stcB, stcF, stcL* and *stcS*) have been identified in the sterigmatocystin cluster of *A. nidulans* and one (*avnA*) has been identified in the *A. parasiticus/A. flavus* aflatoxin cluster (Brown *et al.*, 1996; Keller *et al.*, 1995; Yu *et al.*, 1995;Yu *et al.*, 1997). StcL has been shown to play a role in the desaturation of the bisfuran moiety in the intermediate versicolorin B (Kelkar *et al.*, 1997). *avnA* is involved in the conversion of averantin to averufin and *stcF* appears to be its homologue in sterigmatocystin biosynthesis (Yu *et al.*, 1997).

Other classes of monooxygenases have been identified in polyketide biosynthetic pathways. The sterigmatocystin biosynthetic cluster of *A. nidulans* includes genes similar to chloroperoxidases (*stcC*) and flavin-containing monooxygenases (*stcW*), although the role of the enzymes they encode in relation to this pathway has not been characterised (Brown *et al.*, 1996). Emodinanthrone oxidase, involved in the biosynthesis of emodin, has been purified and characterised as a nonheme $Fe3^{+}$ monooxygenase (Chen *et al.*, 1995). Tetracenomycin (Tcm) F1 monooxygenase, the product of the *tcmH* gene, catalyses the oxidation of the naphthacenone Tcm F1 in the biosynthesis of the anthracycline antibiotic Tcm C in *S. glaucescens* (Shen & Hutchinson, 1993a; Summers *et al.* 1993). Biochemical characterisation of the enzyme identified it as an internal monooxygenase, requiring only O_2 for activity and revealed that this protein lacks the usual monooxygenase prosthetic groups, such as flavin or haem groups, and does not utilise metal ions (Shen & Hutchinson, 1993a). Other enzymes of this class may be involved in anthracycline biosynthesis at oxidation of 12-deoxyaklanonic acid, and genes showing some homology to *tcmH* that may encode this type of enzyme have been cloned from *S. peucetius* and *Streptomyces* sp. C5 (Grimm *et al.*, 1994; Ye *et al.*, 1994).

Bayer-Villiger type oxidative cleavage of one of the aromatic rings is a key reaction in the conversion of an anthraquinone to a benzophenone (Hutchinson & Fujii, 1995). A monooxygenase activity catalysing such a reaction, the questin oxygenase system, has been identified in an *Aspergillus terreus* cell-free extract and partially characterised as consisting of at least two proteins and requiring NADPH, O_2 and possibly electron transport protein(s) (Fujii *et al.*, 1988).

Aspergillus terreus dihydrogeodin oxidase (DHGO) catalyses the phenol oxidative coupling reaction converting dihydrogeodin to (+)-geodin. Polyclonal antibodies raised against the purified protein have been used to obtain a cDNA fragment of the gene for DHGO which was then used to screen genomic and cDNA libraries to obtain full-length

clones (Huang *et al.,* 1995). Analysis of the predicted protein sequence, together with the blue colour of the purified enzyme preparation identified the enzyme as a multicopper blue protein (Huang *et al.*, 1995).

1.5.3 Methylation

Secondary metabolic pathways often involve methylations of hydroxyl or carboxyl groups and polyketides will often undergo more than one methylation during post-PKS tailoring (Hutchinson & Fujii, 1995). The enzymes which catalyse these reactions are known as methyltransferases and the methyl group donor is generally *S*-adenosyl-L-methionine (SAM) (O'Hagan, 1991). In fungi *O*- and *N*-methylations tend to occur towards the end of biosynthetic pathways and are presumably catalysed by enzymes separate from the PKS (O'Hagan, 1991). *C*-methylations probably take place during polyketide assembly and the enzymic activity in these cases may well reside within the PKS, as has been observed in the case of lovastatin biosynthesis (Davis *et al.*,1994; O'Hagan, 1991).

A SAM-dependent methyltransferase which converts sterigmatocystin to *O*-methylsterigmatocystin in the biosynthesis of aflatoxin B1 has been purified from *Aspergillus parasiticus* (Bhatnagar *et al.*, 1988; Keller *et al.*, 1993). This 40 kDa protein is one of three distinct methyltransferases in *A. parasiticus* which will catalyse this conversion (Keller *et al.*, 1993; Yabe *et al.*, 1989). Polyclonal antiserum raised against the purified 40 kDa protein has been used to isolate a full-length *A. parasiticus* cDNA containing the coding region of the gene *omtA* (Yu *et al.*, 1993). The *A. parasiticus* cDNA probe was subsequently used to clone *omtA* from the aflatoxin producer *A. flavus*; this homologue has a high sequence identity (97%) and an identical intron structure (Yu *et al.*, 1995).

Three methyltransferase genes, *tcmNOP,* have been identified within the tetracenomycin C biosynthetic gene cluster (Decker *et al.*, 1993; Summers *et al.*, 1992). The *tcmN* gene product appears to be a multifunctional cyclase-dehydratase-*O*-methyltransferase (Summers *et al.*, 1992). Unusually, TcmP does not exhibit significant predicted amino acid sequence similarity to any other *O*-methyltransferases, including TcmN or TcmO (Decker *et al.*, 1993). SAM-dependent *O*-methyltransferases usually exhibit a highly conserved nucleotide binding motif and the absence of this motif in TcmP implies a different evolutionary origin (Decker *et al.*, 1993).

CHAPTER 2

MATERIALS AND METHODS

2.1 MICROBIAL STRAINS

Escherichia coli

Strain	Genotype	Reference / Supplier
MC1022	araD139, Δ(ara-leu)7697, lacZΔM15, galU, galK, strA	Casadaban & Cohen (1980)
DH5α®	ϕ80dlacZΔM15, recA1, endA1, gyrA96, thi-1, hsdR17 (r_K-, m_K+), supE44, relA1, deoR, Δ(lacZYA-argF)U169	Hanahan (1983)
KW251	supE44, galK2, galT22, metB1, hsdR2, mcrB1, mcrA, [argA81::Tn10], recD1014	Promega

Table 2.1 ***E. coli*** **strains**

Filamentous fungi

The various strains of filamentous fungi used in this study are listed and described in table 4.1.

2.2 CLONING VECTORS

Plasmid

Plasmid	Comment	Reference / Supplier
pUBS1	Derived from Bluescript M13(+/-) (Stratagene)	Dr George Murphy.
pUC18	M13-derived cloning vector	(Vieira & Messing, 1982)
pNEB193	Derived from pUC19	New England Biolabs
pGEM-T	T-overhang vector used for the cloning of PCR products	Promega
pGEX-BHE(E)XSX	Derived from pGEX-4T-2 GST fusion expression vector	Dr Steve Screen

Table 2.2 Plasmid cloning vectors

Phage

The *P. patulum* genomic library (Walsh, 1993) referred to in chapters 3 and 5 of this thesis was constructed in half-site *Xho*I LambdaGEM™-11 vector arms (Promega).

2.3 CHEMICALS

All reagents and buffer ingredients used in this study were of analytical quality (AnalaR® or equivalent standard) or molecular biology grade.

Water was purified by reverse osmosis and ion-exchange using a Labwater RO 100 system (Purite).

Oligonucleotides were synthesised in house by Dr Colin Lazarus.

Isotopically labelled nucleotides were supplied by Amersham.

All reagents and kits were used according to the maufacturer's instructions unless noted otherwise.

2.4 GROWTH MEDIA AND ANTIBIOTICS

Growth media and media ingredients were purchased from Difco or BetaLab.

2.4.1 Growth medium for *E. coli*

Luria-bertani (LB) medium

Bacto-tryptone	10 g
Yeast extract	5 g
NaCl	5 g
H_2O	to 1 litre

The prepared medium was sterilised by autoclaving at 121° C for 15 minutes. For solid medium agar was added to a final concentration of 1.5 % prior to autoclaving.

2.4.2 Growth media for *Penicillium patulum*

Potato dextrose medium and Czapek Dox medium were prepared according to the manufacturer's instructions and sterilised by autoclaving at 121° C for 15 minutes.

2.4.3 Antibiotics

Antibiotic	Solvent	Stock conc.	Working conc.	Sterilisation
Ampicillin	H_2O	100 mg/ml	100 µg/ml	0.2 µm filter
Streptomycin	H_2O	100 mg/ml	100 µg/ml	0.2 µm filter
Tetracycline	ethanol	10 mg/ml	15 µg/ml	none

Table 2.3 Antibiotic stock solutions and working concentrations

2.5 CULTURE CONDITIONS

2.5.1 *E. coli*

Bacterial cultures were generally grown overnight on LB medium at 37° C. Agar plates were inverted during incubation. Liquid cultures were grown in an orbital incubator (Gallenkamp) at 200 rpm.

2.5.2 *P. patulum*

Preparation of a spore suspension

Spores from a single colony were streaked over the surface of a Czapek Dox agar plate and incubated at 27-28° C until the surface of the plate was covered in mature greenish-grey conidia. Five ml of tween saline (tween 80 0.025 %, NaCl 0.8 %) was pipetted onto the surface of each plate and the spores were suspended using a sterile wire scraper. The resulting spore suspension was stored in sealed universal bottles at 4° C.

Mycelial cultures

0.5-1 ml of a spore suspension was used to inoculate 100 ml of Potato Dextrose broth which was incubated at 27-28° C in an orbital incubator shaking at 200 rpm. Mycelium was harvested by filtering through Miracloth (Calbiochem) and growth medium was rinsed away with water. Residual rinse water was removed by gentle squeezing and the harvested mycelium was snap frozen in liquid nitrogen before use or storage at -70° C.

Conidiating cultures

Two 100 ml mycelial cultures were grown as described above and then filtered through sterile No.1 grade filter paper (Whatman) and washed with sterile water. The filter paper with mycelium was transferred under aseptic conditions to a 90 mm petri dish over a layer of glass beads (diameter 2 mm) and 9 ml Czapek Dox medium. Cultures were incubated at 28° C. Aerial hyphae appeared within 30 h of inoculation, mature blue-green conidiospores could be seen 50-75 hours post-inoculation. To harvest conidiating cultures the filter paper was removed from the culture plate and mycelium was scraped from the underside, the culture was squeezed gently to remove excess growth medium, snap frozen in liquid nitrogen and used immediately or stored at -70° C.

2.6 DNA PREPARATION AND PURIFICATION

2.6.1 Plasmid DNA

Plasmid miniprep

To obtain small quantities (1-10 μg) of plasmid DNA the Wizard Miniprep kit (Promega) was used according to the manufacturer's instructions. DNA was eluted in 50 μl water or TE buffer (see appendix) and stored at -20° C.

Plasmid maxiprep

To obtain larger quantities of plasmid DNA of high purity, suitable for applications such as DNA sequencing, the procedure described below was employed. For high copy number plasmids this method typically yielded 200-600 μg DNA.

After overnight incubation in a 100-200 ml culture (under the appropriate antibiotic selection) cells were pelleted by centrifugation at 3650 x *g* for 10 minutes and resuspended in 5 ml of resuspension solution (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 μg/ml RNase A). An equal volume of lysis solution (0.2M NaOH, 1% SDS) was added and mixed by inverting the tube several times until the cell suspension cleared. Five ml of Neutralisation Solution (2.55 M potassium acetate, pH 4.86) was added and mixed into the lysate by inverting the tube several times. Chromosomal DNA and cell debris were pelleted by centrifugation at 50000 x *g* for 10 minutes. The supernatant was decanted to a fresh centrifuge tube and DNA was precipitated by the addition of 0.6-1 volume of ice-cold isopropanol and incubation at -20° C for 15 minutes. Precipitated DNA was pelleted by centrifugation at 50000 x *g* for 10 minutes. The supernatant was aspirated and the pellet resuspended in TE buffer. Plasmid DNA was purified by two rounds of equilibrium centrifugation (Sambrook *et al.*, 1989) in a continuous CsCl-EtBr gradient (TE buffer containing CsCl at a density of 1.56 g/ml and EtBr at 0.4 mg/ml). Centrifugation was carried out using a TL-100 ultracentrifuge (Beckman) with a TLA-100.2 fixed-angle rotor at 100000 rpm for 5-16 hours at 15°C. The plasmid DNA band was visualised using a UVL-21 long wave UV lamp (Ultra-Violet Products Inc.), removed with a hypodermic needle and syringe and diluted in 2.5 volumes of water. CsCl and EtBr were removed from the DNA solution by the addition of 0.6-1 volumes ice-cold isopropanol or 2.5 volumes ice cold ethanol and precipitation at -20° C for 15 minutes followed by centrifugation at 20

000 rpm for 10 minutes to pellet the DNA. The pellet obtained was dried under vacuum and resuspended in 200 µl TE. Sodium acetate was added to 0.3 M and the plasmid DNA was re-precipitated by the addition of 2.5 volumes ice-cold ethanol and incubation at -20° C for 15 minutes. Plasmid DNA was pelleted by centrifugation and the pellet was washed in ice-cold 75% ethanol, dried under vacuum and resuspended in TE buffer.

2.6.2 Phage DNA

DNA was extracted from λ phage particles prepared from plate lysates using a phage absorbent.

Single plaques were transferred into 100 µl of SM buffer (see appendix) and agitated vigorously before being allowed to stand for 30 minutes. Plating cells were added (200 µl of *E. coli* KW251) and incubated at room temperature for 15 minutes. Three ml of top agarose (0.6 % in 10 mM $MgSO_4$) was added and immediately poured onto 90 mm LB agar plates supplemented with tetracycline. Plates were incubated at 37° C without inversion; confluent lysis was observed within 10 h. Phage particles were eluted from the medium into 5 ml of SM buffer, which was pipetted onto the surface, by shaking overnight on a rotary table at 4° C. Cell debris was removed from the buffer by centrifugation at 3650 x *g* for 15 minutes. If the plate lysate was to be stored (at 4° C) rather than used immediately, chloroform was added to a final concentration of 0.3 %.

Pure phage DNA was isolated from the plate lysate using LambdaSorb® (Promega). One hundred µl of this phage adsorbent was added per 10 ml of lysate in a 50 ml centrifuge tube and gently mixed on a MM1 roller-mixer (Luckham) for 30 minutes. Adsorbed phage was pelleted by centrifugation at 3650 x *g* for 20 minutes, resuspended (by pipetting) in 1 ml of SM buffer and transferred to a microcentrifuge tube. After centrifugation at 13000 rpm in a microcentrifuge for 2 minutes, the resulting pellet was washed again in 1 ml of SM buffer before resuspension in 0.5 ml of 10 mM Tris-HCl (pH 8), 10 mM EDTA. Phage DNA was released by heating at 67° C for 5 minutes and the adsorbent was pelleted by centrifugation in a microcentrifuge at 13000 rpm for 5 minutes. The supernatant was transferred to a fresh tube and NaCl was added to a concentration of 500 mM. The phage DNA was purified by two extractions in phenol:chloroform:isoamyl alcohol as described in section 2.6.5 followed by one extraction in chloroform:isoamyl alcohol. DNA was precipitated from the final aqueous phase by ethanol precipitation as described in section 2.6.6. The DNA pellet

obtained was dissolved in 100 µl TE buffer and allowed to dissolve at 4° C overnight before use or storage at 4° C.

2.6.3 Fungal genomic DNA

Approximately 4 g of mycelium was ground to a fine powder under liquid nitrogen using a pre-cooled mortar and pestle. The ground tissue was added to 10 ml of extraction buffer ((10 mM Tris-HCl pH 8, 10 mM EDTA, 0.5% SDS) in a 30-50 ml centrifuge tube, thawed and mixed gently by inversion. Ten ml phenol:chloroform:isoamyl alcohol (25:24:1; Sambrook *et al.*, 1989) were added and mixed gently for 15-30 minutes. The organic and aqueous phases were separated by centrifugation at 3650 x *g* for 20 minutes or 50000 x *g* for 10 minutes and the aqueous layer removed to a to fresh centrifuge tube. This extraction was repeated until the interface between the phases was clear and was followed by an extraction with chloroform:isoamyl alcohol (24:1). DNA was ethanol precipitated from the final aqueous phase, washed in 75% ethanol and re-dissolved in 1 ml TE buffer by gentle shaking overnight on a rotary table. Genomic DNA was purified by two rounds of equilibrium centrifugation in a CsCl gradient (as described in section 2.6.1) and the final pellet of genomic DNA was allowed to dissolve in TE buffer overnight at 4° C. All pipetting of solutions containing high molecular weight DNA was carried out very slowly or using wide-bore pipette tips to minimise shearing.

2.6.4 Purification of DNA fragments from agarose gel slices

Three different methods were used to purify size-fractionated DNA fragments from an agarose gel: the Wizard™ PCR preps DNA purification system (Promega) was used according to the manufacturer's instructions; alternatively the DNA was eluted by centrifugation through a filter, via either the 'Wizard minicolumn' method of Wolff and Hull (1996) or by the method described below:

Elution through glass wool filter

The band of interest was excised, in the minimum possible volume, from an agarose gel which had been cast and run in TAE buffer. The DNA was eluted by centrifugation at approximately 5000 x *g* for 10 minutes through a small amount of glass wool which had been previously siliconised in dimethyldichlorosilane. The DNA dissolved in TAE buffer

thus obtained was diluted with 3-4 volumes of water before use in applications such as the preparation of ^{32}P labelled DNA probes.

2.6.5 Phenol:chloroform extraction

Protein was removed from DNA preparations by extraction in an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The two phases were mixed on a rotary table or by vortexing until a homogeneous emulsion was obtained. Aqueous and phenol phases were separated by centrifugation in a microcentrifuge at 13000 rpm for 10 minutes or at 3650 x *g* for 15 minutes. The aqueous phase was removed to a clean tube and re-extracted in a similar fashion with chloroform:isoamyl alcohol (24:1) to remove lingering traces of phenol.

2.6.6 Ethanol precipitation

The following procedure was routinely used to concentrate and purify DNA solutions. DNA was precipitated by adding sodium acetate to 0.3 M, or ammonium acetate to 2.5 M, followed by 2-2.5 volumes of ice-cold ethanol, mixing and incubating on ice for 15 minutes. Precipitated DNA was recovered by centrifugation at 13000 rpm in a microcentrifuge or at 50000 x *g* for 10-15 minutes in a high speed centrifuge. The resultant pellet was washed in 75 % ethanol to remove contaminating salts and dried under vacuum before redissolving in TE buffer or water.

2.6.7 Analysis of DNA preparations

Genomic and plasmid DNA preparations were generally analysed quantitatively and qualitatively by the following methods. Absorbance (OD) was measured at 260 nm and the value obtained was multiplied by a factor of 50 and the dilution factor to give the DNA concentration in μg/ml. The OD at 280 nm was also measured and the ratio of OD_{260}:OD_{280} could be used to estimate the purity of the DNA preparation (Sambrook *et al.*, 1989). For a more empirical quantification, aliquots of the DNA sample were fractionated on an agarose gel and compared to DNA standards of known molecular weight. To confirm the identity of preparations of plasmid DNA, small amounts were digested with appropriate restriction enzymes and the sizes of the resulting fragments were checked by agarose gel electrophoresis.

2.7 RNA PREPARATION

2.7.1 Total RNA

Approximately 400 mg of fungal tissue were frozen and ground under liquid nitrogen in a precooled pestle and mortar with acid-washed sand added to facilitate grinding. The powdered tissue was transferred to a precooled 2 ml screw-capped microfuge tube. To each tube was added 0.5 ml of extraction buffer (GuHCl 8 M, MES [4-morpholino-ethanol-sulphonic acid] 20 mM, EDTA 20 mM, 2-mercaptoethanol 0.35 % v/v) and 0.5 ml of acid buffered (pH 6) phenol:chloroform:isoamyl alcohol (25:24:1) and the mixture was vortexed until a stable emulsion was formed. The aqueous and organic phases were separated by centrifugation at 13000 rpm in a microcentrifuge for 10 minutes and the aqueous phase removed to a precooled 1.5 ml microfuge tube. The aqueous phase obtained was re-extracted with phenol:chloroform:isoamyl alcohol until the interface between the two phases was clear after centrifugation. Nucleic acids were precipitated from the final aqueous phase by the addition of acetic acid (to 0.17 M) and ice cold ethanol (to 40 %) and overnight incubation at -20° C. Precipitated nucleic acids were pelleted by centrifugation at 13000 rpm for 15 minutes in a microcentrifuge and the pellet was washed twice by vortexing in 400 µl sodium acetate (3 M, pH 5.5) at 4° C to remove low molecular weight RNA species and polysaccharides. The salt was removed from the washed pellet by a final wash in 400 µl ice-cold 70 % ethanol. The pellet was dissolved in 60 µl H_2O by heating at 95° C for 2 minutes, vortexing and rapid cooling on ice. Quantity and quality of the RNA prepared was assessed by measurement and comparison of absorbance at 260 and 280 nm, followed by electrophoresis and EtBr staining of approximately 1 µg on a 1 % agarose gel in 0.5 x TBE buffer (see section 2.12.1). An RNase inhibitor (36 U of RNAguard [Pharmacia]) was added and the sample was stored at -70 ° C.

2.7.2 Isolation of mRNA

Poly $(A)^+$ RNA was isolated from *P. patulum* total RNA using the PolyA Spin™ mRNA isolation kit [New England Biolabs] according to the manufacturer's instructions. The final pellet of poly $(A)^+$ RNA obtained was dissolved in 50 µl water by heating to 95° C for 2 minutes, vortexing and cooling on ice for 5 minutes before use or storage at -70° C. The mRNA content was quantified by staining a 1 µl aliquot with an equal volume of EtBr (1

μg/ml) and comparison of the fluorescence of the sample droplet with that of a series of RNA standards of known concentrations.

2.8 cDNA SYNTHESIS

First strand cDNA was synthesised from 60-100 ng of poly $(A)^+$ RNA using the Ready-To-Go™ T-primed First-Strand kit (Pharmacia) according to the manufacturer's instructions.

2.9 RESTRICTION ENZYME DIGESTION

Restriction enzymes were purchased from Gibco-BRL or Boehringer-Mannheim and reactions generally used the buffer and temperature recommended by the manufacturer. For double digests involving enzymes with different buffer requirements Multi-Core™ buffer (Promega) was used. To avoid inhibition of digestion by glycerol in the enzyme storage buffer, the volume of enzyme added comprised less than 10 % of the total reaction volume in each case. Restriction enzymes were heat inactivated, where appropriate, by incubation at 70° C for 15 minutes.

2.10 DNA CLONING

2.10.1 Preparation of competent *E. coli*

A starter culture of *E. coli* strain MC1022 or DH5α® (table 2.1), inoculated from a single colony, was grown in 10 ml LB broth at 37° C in an orbital shaker. After overnight incubation, 1 ml was added to 100 ml of LB broth and incubated for up to 3 hours at 37° C on an orbital shaker, until the cells reached mid-log growth phase, i.e. until OD_{600nm} for the culture was approximately 0.6. Cells were cooled on ice for 15 minutes and then pelleted by centrifugation at 3650 x *g* for 10 minutes. The supernatant was removed and the cells resuspended in 50 ml ice-cold 0.1 M magnesium chloride. The cells were pelleted again and resuspended in 50 ml 0.1 M calcium chloride. After 30 minutes on ice the cells were pelleted and resuspended in 6 ml ice-cold 0.1 M calcium chloride. Cells were kept on ice for immediate use and remained viable for 24 hours. For long term storage of competent cells, sterile glycerol was added to a final concentration of 15 % and 1 ml aliquots were snap-frozen in liquid nitrogen prior to storage at -70° C.

2.10.2 Plasmid vector and insert preparation

Standard vector preparation involved digestion of 2.5 μg of plasmid DNA with the appropriate restriction enzyme in a 25 μl reaction volume, followed by heat inactivation of the enzyme. The reaction mixture was then diluted to 100 μl with TE buffer before dephosphorylation of the DNA by addition of 6 units of alkaline phosphatase and incubation at 37° C for 15 minutes. The alkaline phosphatase was inactivated by addition of nitrilotriacetic acid to 20 mM and incubation at 70° C for 10 minutes. DNA was then ethanol precipitated from the solution and resuspended in 20 μl TE buffer to give a DNA concentration of approximately 100 ng/μl.

Insert DNA was digested with the appropriate restriction enzymes which were then heat inactivated. Ammonium acetate was added to 2.5 M and the DNA was ethanol precipitated and resuspended in an appropriate volume of TE buffer or water prior to ligation. Where the insert fragment desired was one of several fragments obtained on restriction digestion, it was isolated from an agarose gel as described in section 2.12.1.

2.10.3 Ligation reactions

Ligation of the DNA fragment to be cloned into the plasmid vector was achieved either by use of the protocol described below or by using the Rapid Ligation Kit (Boehringer Mannheim). A molar ratio of vector:insert between 1:1 and 1:3 was used in most cases. 100 ng of the digested vector was ligated to an appropriate amount of insert DNA in a total volume of 10 μl using 1 unit of T4 DNA ligase and the ligation buffer supplied by the manufacturer (Gibco-BRL or Boehringer-Mannheim). A control reaction was prepared without insert DNA to allow the frequency of vector religation to be assessed. The reaction mixture was incubated at 12° C overnight and then 30 μl of TE buffer was added and the ligase enzyme heat-inactivated by incubation at 70° C for 10 minutes.

2.10.4 Transformation of competent *E. coli*

Half of the inactivated ligation mix was added to 200 μl of competent cells in precooled microfuge tubes and incubated on ice for 30 minutes, followed by a 2 minutes incubation at 37° C. 1 ml of LB broth was added to each transformation mixture and after gentle mixing the cells were incubated for 1 hour at 37° C. Aliquots (20, 60 and 200μl) were spread onto LB agar containing the appropriate selective antibiotic and, for blue-white colour screening with vectors capable of α-complementation, IPTG (0.1 mM) and X-gal (5-bromo-4-chloro-

3-indolyl ß-*D*-galactosidase, 32.5 µg/ml). To enable transformation frequencies to be assessed, a 200 µl aliquot of cells was transformed with 1 ng of uncut plasmid as a positive control. The transformed cells were screened for the required recombinants either by plasmid miniprep and restriction enzyme digestion or by single colony PCR (see below).

2.10.5 Single colony PCR

This technique was particularly useful when the vector used for cloning did not allow blue white screening and large numbers of transformant colonies had to be checked for the presence of recombinant plasmid. A small amount of each colony was removed with a sterile toothpick, resuspended in 20 µl lysis buffer (TE buffer with proteinase K 50 µg/ml) and incubated for 15 minutes at 55° C to lyse the cells. The proteinase K was inactivated by incubation for 15 minutes at 80° C and the lysates were cooled on ice. Cell debris was pelleted by centrifugation at 13000 rpm for 3 minutes in a microcentrifuge and 5-15 µl of the supernatant was used as template in a 25 µl Supertaq PCR (see section 2.15).

2.11 PHAGE LIBRARY SCREENING

2.11.1 Preparation of *E. coli* plating cells

A single colony of KW251 (table 2.1) was inoculated into 100 ml LB with tetracycline and incubated overnight in an orbital incubator at 37° C, 200 rpm. The cells were pelleted by centrifugation at 3650 x *g* for 10 minutes and resuspended in 50 ml ice cold $MgSO_4$ (10 mM). The cells were pelleted again, resuspended in 25 ml ice-cold $MgSO_4$ (10 mM) and stored on ice until use (for a maximum of 1 month).

2.11.2 Plating phage

Aliquots of phage suspension were mixed with 200 µl plating cells and incubated at room temperature for 15 minutes in a sterile 5 ml tube. Three ml of molten top agarose (0.6 % in 10 mM $MgSO_4$) were added and quickly poured onto LB agar plates containing tetracycline. After the top agar had been allowed to set at room temperature, plates were inverted and incubated at 37° C overnight.

2.11.3 Transfer to nitrocellulose membrane and screening of phage DNA

Nitrocellulose filters (0.45 μm, Schleicher & Schuell) were laid onto phage plates and marked using a needle and Indian ink to facilitate subsequent orientation. The filter was left in place for 1 minute to allow transfer of phage particles before being peeled off and transferred 'phage side up' to a pad of 3MM Chr paper (Whatman) soaked in 0.5 M NaOH. Then the filter was transferred to a pad soaked in a second denaturing buffer (0.5 M NaOH, 1.5 M NaCl) for 20 seconds and then to a pad soaked in neutralising buffer (0.5 M Tris-HCl pH 7.5, 1.5 M NaCl) for a further 20 seconds. Finally the filter was totally immersed and washed in the neutralising solution for 20 seconds and then washed for 1 minute in 3 x SSC. Filters were air dried and then baked at 80° C for 30 minutes. If duplicate transfers were made, the second filter was left in place for 5 minutes before processing.

After hybridisation of labelled probes to the filter, the autoradiograph was aligned with the original plate and plaques identified as containing the clone(s) of interest were removed in agar plugs with a sterile pipette tip. The phage were allowed to diffuse into 1 ml of SM buffer containing 1.5% chloroform. Phage suspensions were stored at 4° C. To re-screen the picked plaque, dilutions of the phage suspension were made in SM buffer and aliquots re-plated.

2.12 GEL ELECTROPHORESIS

2.12.1 Agarose gel electrophoresis of DNA

Horizontal gel electrophoresis equipment supplied by various manufacturers was used to cast and run agarose gels. Agarose gels were prepared at various concentrations to maximise the resolution of the DNA fragments under investigation. TAE buffer (Tris-acetate 40 mM, EDTA 1 mM) was used when gels were run overnight or when a DNA fragment was to be purified from the gel; otherwise a 0.5 x TBE (Tris-borate 45 mM, EDTA 1 mM) buffering system was used (Sambrook *et al.*, 1989). Agarose gel loading buffer (10x concentration: bromophenol blue 0.25 % w/v, glycerol 50 % v/v) was added to the DNA sample prior to loading onto the gel. Gels were run at a constant current appropriate for the size of the DNA fragments to be resolved. DNA fragments were visualised on a TM-36 UV transilluminator (Ultra-Violet Products Inc.), after the gel had been stained in an EtBr (approximately 0.5 μg/ml) for 5 minutes. Gel images were routinely recorded using a video camera linked to a Video Copy Processor (Mitsubishi)

with thermal printer; presentation quality photographs were taken using a direct screen instant camera and 667 black and white film (Polaroid).

2.12.2 Agarose gel electrophoresis of RNA

For purposes such as visual assessment of total RNA quality and quantity, a 1.25 % agarose gel cast and run in a 0.5 x TBE buffer (see above) was used to fractionate RNA. For northern blotting total RNA was fractionated on a formaldehyde denaturing gel. The 0.75% agarose gel was cast in MOPS buffer (3-[*N*-morpholino]propanesulphonic acid 20 μM, sodium acetate 8 mM, EDTA 1 mM) with formaldehyde added to 1.75 M and run in MOPS buffer with 1.2 M formaldehyde added at 25 V, 25 mA. RNA samples were freeze dried and dissolved in 10 μl RNA loading buffer (1 x MOPS buffer, 6.5 % formaldehyde, 50 % formamide, 1 x agarose gel loading buffer [see section 2.12.1]), then heated to 65° C for 5 minutes and rapidly cooled on ice before loading. Samples were run alongside 0.24-9.5 kb RNA molecular size standards (Gibco-BRL). This lane was cut from the gel prior to blotting, stained with EtBr and used to estimate the sizes of bands in the sample lanes.

2.12.3 Polyacrylamide gel electrophoresis of DNA sequencing reaction products

The radiolabelled products of manual sequencing reactions (see section 2.16.1) were fractionated by electrophoresis through urea (7.7 M)-polyacrylamide (6 %) wedge gels (0.4-0.8 mm) in 3:1 TBE buffer (Tris-HCl 135 mM, Boric acid 45 mM, EDTA 2.5 mM). The gels were run in an IBI vertical apparatus attached to an E734 power supply unit (Consort). Gels were prewarmed for one hour at 60W with the shark's-tooth well-forming comb in place and then urea was rinsed from the wells with 3:1 TBE buffer. A 2-5 μl aliquot of a sequencing reaction was loaded in each lane and the gel was run at 55-60 W until the bromophenol blue dye front reached the bottom of the gel. The gel was fixed in methanol (10 %) acetic acid (10%) for 30 minutes, the top 17 cm of the gel was removed and the remainder was transferred to a 3MM Chr paper (Whatman) support and dried on a slab dryer at 80° C for at least one hour. The dried gel was autoradiographed by exposure to Hyperfilm-MP (Amersham) or Biomax MR-1 (Kodak) film.

2.12.4 SDS-polyacrylamide gel electrophoresis of proteins

Protein samples were fractionated using a 10 % polyacrylamide separating gel and a 4 % stacking gel made with 37.5:1 acrylamide:bisacrylamide solution. Polymerisation was

catalysed by the addition of ammonium persulphate to 0.1 % and NNN'N'-tetramethylethylenediamine to 0.01 %. The gel (1 mm thickness) was cast and run in a Mini Protean II™ apparatus (Bio-Rad) using Tris/glycine/SDS running buffer (National Diagnostics). Electrophoresis was carried out at 200 V until the bromophenol blue dye in the sample loading buffer reached the bottom of the gel. The gel was stained overnight in a solution of 1 % Coomassie blue in destain solution (methanol 40 %, acetic acid 10 %) at room temperature with gentle shaking and then destained for several hours with regular changes of the destain solution.

2.13 TRANSFER OF DNA AND RNA TO NYLON MEMBRANES

2.13.1 Southern blotting

DNA was transferred to Zeta-Probe® nylon membrane (Bio-Rad) by alkaline capillary transfer for 5-8 h using 0.4 M NaOH as transfer buffer and the protocol supplied by the manufacturer of the membrane. Membranes were dried on 3MM Chr paper (Whatman) at 80° C for 30 minutes before use. To improve transfer of large DNA fragments (greater than 4 kb) the gel was soaked in 0.25 M HCl for 15 minutes to partially depurinate the DNA and then briefly rinsed in water prior to blotting.

2.13.2 Northern blotting

RNA was transferred to Zeta-Probe® membrane by alkaline capillary transfer in 50 mM NaOH. The transfer was allowed to proceed for 3-5 h and then the membrane was rinsed in 2 x SSC, blotted dry on 3MM Chr paper (Whatman) and baked at 80° C for 30 minutes before use.

2.14 PROBING SOUTHERN AND NORTHERN BLOTS

2.14.1 Preparation of radioactive DNA probes

Inserts were excised from recombinant plasmids by restriction digestion and isolated from agarose gels prior to labelling. PCR products were purified using the HighPure (Boehringer-Mannheim) or Wizard™ PCR preps (Promega) DNA purification systems. DNA fragments were labelled with [α-^{32}P]dCTP using either the Rediprime kit (Amersham) or Ready To Go™ DNA-Labelling (-dCTP) reaction mixes (Pharmacia).

Unincorporated nucleotides were removed from probes using Nick columns (Pharmacia) according to the manufacturer's instructions. Probes were eluted from Nick columns in 400 µl water, denatured by boiling for 5 minutes and then cooled on ice before immediate use or storage at -20° C.

2.14.2 Preparation of non-radioactive RNA probes

Digoxygenin (DIG) labelled RNA probes were prepared by *in-vitro* transcription from the SP6 and T7 promoters of LambdaGEM™-11 phage clones using DIG RNA labelling mix (Boehringer Mannheim).

2.14.3 Hybridisation of labelled probes to Southern or northern blots

The following protocol, adapted from that supplied with the Zeta-Probe® nylon membrane (Bio-Rad) was used for all hybridisations of ^{32}P-labelled DNA probes to Southern or northern blots. The membrane was prehybridised in 40 ml of hybridisation buffer (0.25 M phosphate buffer pH 7.2 [Sambrook *et al*, 1989], 7 % SDS) for 5 minutes at the hybridisation temperature and then this buffer was removed and replaced with 10 ml of fresh buffer to which the probe was added. Hybridisation reactions were incubated overnight in an HB-1 hybridising oven (Techne). The hybridisation buffer was then discarded and the membrane rinsed briefly in wash buffer (2 x SSC, 0.1 % SDS) followed by two 15 minutes washes in the same buffer. All hybridisation and washing steps were performed at 65° C except where noted. Finally, the membrane was blotted dry and wrapped in Saran™ wrap (Dow) before autoradiography.

2.14.4 Autoradiography of Southern and northern blots.

Post-hybridisation and washing, membranes were exposed to Hyperfilm MP film (Amersham) in a cassette with intensifying screen at -70° C. Films were developed using LX-24 x-ray developer (Kodak) and Hypam rapid fixer (Ilford).

2.14.5 Stripping filters

Probes were stripped from blots by either the alkaline denaturation or SDS/SSC wash methods described in the protocol supplied with the Zeta-Probe® membrane (Bio-Rad). Stripped membranes were autoradiographed to ensure that the probe had been removed.

2.15 POLYMERASE CHAIN REACTION.

For routine PCR, where a high fidelity amplification was not required, the non proof-reading DNA polymerase Super Taq (HT Biotechnologies) was used in a 25 µl reaction using the buffer supplied under the following standard conditions. Template DNA concentrations were approximately 4-40 ng/µl for genomic DNA, 0.04-0.4 ng/µl for plasmid DNA. Primers were both at 1 µM, dNTPs were each at 0.2 mM and 0.5 U of enzyme was used in a 25 µl reaction. Thermal cycling parameters were: (94° C, 3 minutes) x 1, (94° C, 1 minute; 55° C, 1 minute; 72° C, 3 minutes) x 32, (72° C, 10 minutes) x 1.

Where the PCR product was to be cloned and sequenced, the Expand™ proof-reading enzyme mixture (Boehringer-Mannheim) was used for amplification. In a typical 25 µl reaction, both primers were used at 0.3 µM, dNTPs were each at 0.2 mM and 0.7 U of enzyme was used. Template DNA concentrations were as described for non proof-reading PCR. Standard thermal cycling parameters were as follows. The initial denaturation stage was at 94° C for 2 minutes and this was followed by 10 cycles of amplification (94° C, 15 seconds; 55° C, 30 seconds; 72° C, 45 seconds). Then the extension (72° C) stage was increased in duration by 15" every 4 cycles for a further 20 cycles of amplification. Finally the reaction was incubated at 72° C for 6'.

2.16 DNA SEQUENCING

2.16.1 Manual sequencing

DNA fragments cloned into plasmid vectors were sequenced using a chain-termination method (Sanger *et al.* 1977). Sequencing reactions were performed using a T7 DNA polymerase sequencing kit (Pharmacia), according to the instructions supplied. Universal (M13) or gene-specific primers were diluted to a concentration of 5 ng/ml before use. Reaction mixtures were labelled with [^{35}-S] dATP and fractionated on a 6% polyacrylamide wedge gel (see section 2.12.3).

2.16.2 Automated fluorescent sequencing

Dye terminator cycle sequencing was performed by Helen Everest (University of Bristol) using an ABI-377 DNA sequencer (Perkin Elmer).

2.16.3 Analysis and manipulation of sequence data

The DNASIS (Pharmacia) and University of Wisconsin GCG (Devereux *et al.*, 1984) software packages were used to analyse and manipulate the sequence data obtained.

2.17 EXPRESSION OF RECOMBINANT PROTEIN IN *E. COLI*

GST-fusion constructs were expressed in small-scale (5 ml) cultures of *E. coli* DH5α (table 2.1). A single colony carrying the appropriate plasmid was inoculated into 5 ml of LB broth with ampicillin and incubated overnight at 37° C with shaking at 200 rpm. From this starter culture, 0.5 ml was inoculated into a fresh 5 ml of LB with ampicillin and incubated for 1 h at 37° C with shaking at 200 rpm. A sample of 1 ml (t=0) was removed and processed as described below. The culture was then induced by the addition of IPTG to 1 mM and incubated for a further 3 h, when a second 1 ml sample (t=3) was removed. The samples were pelleted by centrifugation for 1 minutes at 13000 rpm in a microcentrifuge and washed in 300 µl PBS (0.8 % NaCl, 0.02 % KCl, 0.144 % Na_2HPO_4, 0.024 % KH_2PO_4, pH 7.4). After washing the cells were resuspended in 50 µl TE buffer and 50 µl of 2 x sample loading buffer (0.125 M Tris-HCl, pH 6.8; 20% glycerol; 4 % SDS; 10 % ß-mercaptoethanol; 0.04 % bromophenol blue) was added. The samples were boiled for 3 minutes prior to fractionation of an aliquot by SDS-polyacrylamide gel electrophoresis (see section 2.12.4). The remainder was stored at -20° C.

Appendix 1

Common Buffers

TE buffer

10 mM	Tris-HCl, pH 8.0
1 mM	EDTA

Sterilised by autoclaving.

SSC

150 mM	NaCl
15 mM	sodium citrate

Adjust pH to 7.0 with NaOH.

SM buffer

0.01 %	gelatin
50 mM	Tris-HCl, pH 7.5
100 mM	NaCl
8 mM	$MgSO_4$

Sterilised by autoclaving.

CHAPTER THREE

CLONING PATULIN BIOSYNTHESIS GENES

3.1 INTRODUCTION

3.1.1 Isolation of patulin

Patulin (4-hydroxy-4h-furol [3,2-c]pyran-2(6H)-one) is a small lactone formed via the tetraketide route. It was initially isolated from filtrates of *Penicillium claviforme* cultures and was found to possess powerful antibacterial properties (Chain *et al.*, 1942). After promising early trials, including clinical trials, it was found to be too toxic for therapeutic use (Friedman, 1990).

Patulin has been isolated from a number of species of the genera *Penicillium, Aspergillus, Paecilomyces* and *Byssochlamys* (table 3.1).

Penicillium

P. expansum (synonym *P. leucopus*), *P. patulum* Bainier (synonyms *P. urticae, P. griseofulvum*), *P. claviforme, P. equinum* (synonym *P. terrestre), P. novae-zeelandiae, P. lapidosum, P. granulatum* (synonym *P. divergens*), *P. lanosum, P. melinii, P. cylopium, P.cyaneo-fulvum, P. roqueforti.*

Aspergillus

A. clavatus, A. giganteus, A. terreus.

Paecilomyces

P. varioti

Byssochlamys

B. nivea, B. fulva.

Table 3.1 Patulin producing organisms, as reviewed by Betina (1989)

3.1.2 Biological activity

Patulin has been clearly shown to be a potent mycotoxin and a mutagen; evidence of its proposed carcinogenicity and teratogenicity has so far been less conclusive (Friedman, 1990; Hopkins, 1993). Patulin presents a risk to humans and livestock by its presence in various foodstuffs contaminated by patulin-producing moulds, the principal examples being rotted apples and apple products (Friedman 1990).

3.1.3 Patulin as a secondary metabolite

Patulin is a secondary metabolite as defined by Hunter (1992): it arises from a biochemical pathway which is not necessary for growth or reproduction of the producing organism, but which can be demonstrated biochemically. Fungal cultures, particularly under controlled

laboratory conditions, may exhibit two distinct phases of growth: the tropophase, during which replicatory growth takes place and the idiophase, or stationary growth phase, during which secondary metabolism occurs (Bu'Lock, 1980). From this metabolic phasing comes the term idiolites, often used to describe the secondary metabolites. Patulin biosynthesis has been shown to conform to this model (Grootwassink & Gaucher, 1980).

A possible role for secondary metabolites in *Penicillium* sporulation has been suggested, based on the fact that submerged-culture sporulation occurs during the idiophase (Sekiguchi & Gaucher, 1977). Mutant strains of *P. urticae* blocked at various stages of the patulin pathway do not show a reduction in conidiogenesis and therefore patulin biosynthesis does not appear to be a prerequisite to sporulation, although an indirect relationship between the two processes has been proposed (Sekiguchi & Gaucher, 1977).

3.1.4 Patulin biosynthesis in *P. patulum.*

In general, the enzymes involved in secondary metabolism have proved difficult to isolate in an active form (Bennett, 1995; Dutton, 1988; Robinson, 1991). However, in terms of its intermediates, enzymology and blocked mutants, the patulin biosynthetic pathway must be one of the best characterised in fungal secondary metabolism (Betina, 1989, 1995). As currently understood, the pathway from acetyl- and malonyl-CoA to patulin in *Penicillium patulum* involves at least ten discrete steps and possesses a number of side-branches (see figure 3.1). Nine of the pathway-specific enzymes have been characterised to date and the genes for two of these enzymes have been cloned (see table 3.2).The regulation of patulin biosynthesis has also been studied extensively, as described below.

Secondary metabolic processes often show a requirement for a single trace element at concentrations which will not affect vegetative growth. A specific manganese requirement for patulin production has been demonstrated and implicated in a transcriptional control mechanism that influences the co-ordinate appearance of enzymes in the post-6-MSA section of the pathway (Scott *et al.,* 1986a,b).

Patulin production is known to be initiated in response to nutrient nitrogen limitation (Grootwassink & Gaucher, 1980). Production of other fungal polyketides, such as the mycotoxin alternariol and the antiprotozoal bikaverins (Orvehed *et al.*, 1988; Bu'Lock *et al.,* 1974) is also known to be inhibited by nutrient nitrogen, although the biosynthesis of aflatoxin and related metabolites does not appear to be regulated in this manner (Coupland and Niehaus, 1987). This nitrogen metabolite repression is "the most common and

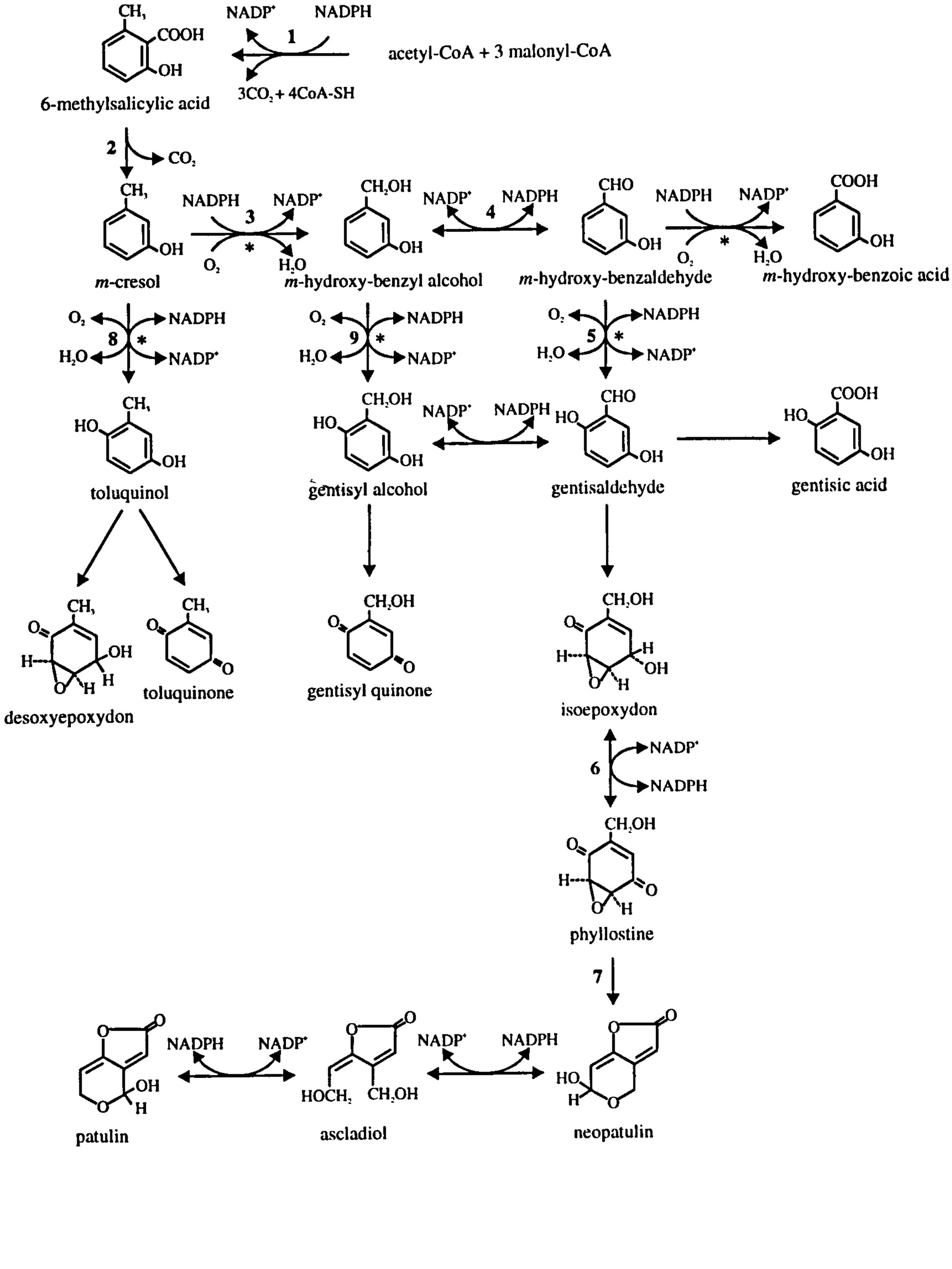


Figure 3.1. The pathway of patulin biosynthesis in *Penicillium patulum.*

Numbered reactions are catalysed by enzymes listed in table 3.2. Reactions potentially catalysed by cytochrome P450 enzymes are indicated with an asterisk.

Adapted from Gaucher *et al.*, 1981.

	Enzyme	Characterised	Purified	Gene Cloned	Reference
1	6-methylsalicylic acid synthase	yes	yes	yes	Schorr et al., 1994; Spencer & Jordan, 1992; Dimroth, Walter & Lynen, 1970; Beck et al., 1990
2	6-methylsalicylic acid decarboxylase	yes	no	no	Light & Vogel, 1975
3	m-cresol methyl hydroxylase	yes	no	no	Murphy et al., 1974
4	m-hydroxybenzyl alcohol dehydrogenase	yes	yes	no	Scott et al., 1986
5	m-hydroxybenzaldehyde hydroxylase	yes	no	no	Murphy & Lynen, 1975
6	isoepoxydon dehydrogenase	yes	yes	yes	Sekiguchi & Gaucher, 1979a,b; GM Gaucher pers. comm.
7	neopatulin synthase	yes	yes	no	GM Gaucher pers. comm.
8	m-cresol 6-hydroxylase	yes	no	no	Murphy et al., 1974
9	m-hydroxybenzyl alcohol 6-hydroxylase	yes	no	no	Murphy & Lynen, 1975

Table 3.2 Patulin biosynthetic enzymes

effective negative control of secondary metabolite biosynthesis" according to Rollins and Gaucher (1994), who showed that ammonium ions, the most repressive nitrogen source, caused a substantial decrease in specific activity of the marker enzyme *m*-hydroxybenzylalcohol dehydrogenase.

Gaucher *et al.* (1981) examined the initiation and longevity of patulin biosynthesis. They found that under defined conditions of nitrogen exhaustion the first enzyme in the pathway, MSAS, appeared at the cessation of replicatory growth. Later enzymes in the pathway appeared together after about another 4 hours, due to a requirement for a co-ordinate induction of these enzymes by early pathway metabolites. It was also observed that the appearance of the pathway enzymes need not be in the same order as their position in the pathway. These rapid enzyme appearances are due to *de novo* transcription and translation of the enzyme genes (Grootwassink & Gaucher, 1980). The appearance of pathway enzymes was transient with MSAS having the shortest half-life of the three enzymes examined. It has been suggested that the degradation of these enzymes may be mediated by the intracellular proteinases which appear during the idiophase (Rollins & Gaucher, 1994).

3.1.5 Genetics of patulin biosynthesis

In contrast to enzymology, elucidation of the genetics of patulin biosynthesis is far from complete. Only two of the structural genes involved, those encoding MSAS (Beck *et al.*, 1990) and IDH (G.M. Gaucher, personal communication), have been cloned and characterised to date. The *MSAS* gene of *P. patulum* was the first fungal PKS gene to be cloned and subsequently a homologue of this gene has been cloned from *Aspergillus terreus* (Fujii *et al.*, 1996; Pazoutova, 1995).

The *nrfA* gene encoding the nitrogen regulatory protein mediating the nitrogen catabolite repression described above has been cloned and sequenced (Ellis, 1996). NRFA appears to be homologous to other fungal GATA-binding proteins of the zinc finger type such as the global nitrogen regulatory factors NIT2 from *Neurospora crassa* (Fu & Marzluf, 1990) and AREA from *Aspergillus nidulans* (Kudla *et al.*, 1990).

3.1.6 Clustering of fungal secondary metabolic pathway genes

As recently as 1987 it was stated that "Structural genes of related function are not often linked with one another in filamentous fungi but instead are distributed, almost at random,

throughout the fungal genome" (Gurr *et al.*, 1987). Only a handful of primary metabolic pathway genes disobey this rule; all are involved in nutrient utilisation pathways, (Keller & Hohn, 1997). An example is the cluster of genes encoding regulatory and structural proteins for quinate catabolism in *Neurospora crassa* (Giles *et al.*, 1985). However it has subsequently become apparent that this is not the case for secondary metabolic pathways, the structural and regulatory genes for which tend to be linked in fungal systems (reviewed by Bennet, 1995; Keller & Hohn, 1997). Gene clustering has been demonstrated for the biosynthetic pathways producing the β-lactam antibiotics penicillin and cephalosporin in *Aspergillus nidulans*, *Penicillium chrysogenum* and *Cephalosporium acremonium* (Mathison *et al.*, 1993; Turner, 1994; Smith *et al.*, 1990) and has facilitated the rapid cloning of genes from these pathways. The biosynthetic genes for the mycotoxins: aflatoxin (Yu *et al.*, 1995), sterigmatocystin (Keller & Adams, 1995) and trichothecene (Hohn *et al.*, 1993; McCormick *et al.*, 1996) are also linked. Genes involved in the biosynthesis of the polyketide melanins of filamentous fungi have been found to be clustered or closely linked in *Alternaria alternata, Cochliobolus heterostrophus* and *Cochliobolus miyabeanus* (Kimura & Tsuge, 1993; Tanaka, Kubo & Tsuda, 1991; Kubo *et al.*, 1989), although no linkage has been observed in *Colletotrichum lagenarium* or *Magnaporthe grisea* (Kubo *et al.*, 1991; Chumley & Valent, 1990).

For obvious reasons this clustering should greatly facilitate cloning and manipulation of secondary metabolic pathway genes.

3.1.7 Evolution of gene clustering in filamentous fungi

The high level of sequence similarity between some fungal and actinomycete β-lactam pathway genes, together with the lack of introns in the fungal genes, has been explained by a proposed horizontal transfer of these genes into fungi from β-lactam producing streptomycetes (Bennet, 1995; Cohen *et al.*, 1990). Such an event, occurring close to the divergence between Gram-positive and Gram-negative bacteria 1.0-1.5 billion years ago, could certainly explain the clustering of these genes. However, this particular horizontal transfer has been refuted by Smith, Feng and Doolittle (1992) who claimed that the phylogenetic tree for isopenicillin N synthetase best supports a conventional evolutionary descent for this gene. Some genes in ß-lactam biosynthetic clusters do contain introns and for this reason are unlikely to be bacterial in origin, but rather have been 'recruited' from

the fungal genome (Mathison *et al.*, 1993). The fact that these recruited genes are closely linked to the others suggests some functional importance for the clustering.

An alternative explanation for the clustering observed in fungal secondary metabolic pathway genes is that it may play a role in the regulation of their expression (Hohn *et al.*, 1993). According to Keller & Hohn (1997) the limited evidence available suggests that when individual genes from a particular cluster are ectopically integrated into the genome their expression does not differ significantly from genes within the cluster, making such a role unlikely. However there is at least one significant exception to this rule: when a reporter construct using the promoter from the gene *ver-1*, part of the aflatoxin biosynthesis cluster in *A. parasiticus*, was integrated at another genomic locus a greatly reduced level of expression was reported (Liang *et al.*, 1997). The authors suggested that this "positional effect" may be due to the topological structure of the chromatin, local *cis*-acting elements or location-specific *trans*-acting mechanisms.

3.1.8 This project

Information about the molecular genetics of patulin biosynthesis is of interest as the biochemistry and enzymology of this pathway have been studied as a model for fungal secondary metabolism. In addition, such information may facilitate the development of control measures for this mycotoxin, such as competitive exclusion of toxigenic strains in the field by genetically engineered, stably non-toxigenic organisms (Linz & Pestka, 1992) or PCR-based post-harvest screening for potentially toxigenic organisms contaminating stored crops (Geisen, 1996).

Working from the hypothesis that patulin pathway genes may be clustered on the chromosome in *P. patulum*, the aim of this project was to examine the chromosomal regions on either side of *MSAS* for open reading frames likely to be required for patulin biosynthesis.

3.2 RESULTS

3.2.1 Chromosome walking

A *P. patulum* genomic library, constructed using λGEM-11 *Xho*I half-site phage arms (Promega), had previously been screened with the *MSAS* condensing and reducing domain probes CON1 and RED3 (Walsh, 1993). Twenty-two positive clones hybridised to one or both probes and plaques corresponding to each of these clones were investigated further to see if any overlapped *MSAS* at the 5' or 3' end.

1 µl aliquots of 10^{-2} dilutions from each picked plaque suspension were dotted onto lawns of *E. coli* KW251 plating bacteria in duplicate and incubated overnight at 37° C. Duplicate plaque lifts were taken from each plate and these were probed using ^{32}P-labelled 5'MS, CON1, RED3 and 3'MS restriction fragments (figure 3.2). Several clones hybridised to all the probes except for 5'MS or 3'MS and could therefore potentially overlap *MSAS*; four of these were assessed further using two sets of PCR primers, shown in figure 3.2. FORP and RCON amplify a fragment from the 5' end of *MSAS* and gave a product of the expected size with clones 13 and 18 as template, FREV and RREV amplify from the 3' end and gave a product of the expected size with clones 3 and 6 as template (figure 3.3). The results confirmed that clones 3 and 6 overlap *MSAS* at the 3' end, while clones 13 and 18 overlap at the 5' end. Clones 6 and 18 were plaque purified by rescreening using the 5'MS and 3'MS probes. Two rounds of purification gave a pure suspension of phage 18α1 (5') while a single round of purification gave pure phage 6α (3'). Phage DNA was prepared from each clone by the plate lysate - Lambdasorb™ method.

3.2.2 Restriction mapping of clones 6α and 18α1.

Approximately 200 ng of DNA was digested with the restriction enzymes *Sal*I and *Hin*dIII for 2 hours at 37° C. The products of these digestions were fractionated on an agarose gel and restriction fragment lengths were ascertained by comparison with DNA size markers (figure 3.4, table 3.3).

Digested DNA was blotted onto a nylon membrane and ^{32}P-labelled *MSAS* probes together with oligonucleotides SP6-20 (CCATTTAGGTGACACTATAG) and T7-24 (ATTGTAATACGACTCACTATAGGG), designed to bind to the RNA polymerase promoters of λGEM-11, were hybridised to the blotted DNA. The data obtained (table 3.4)

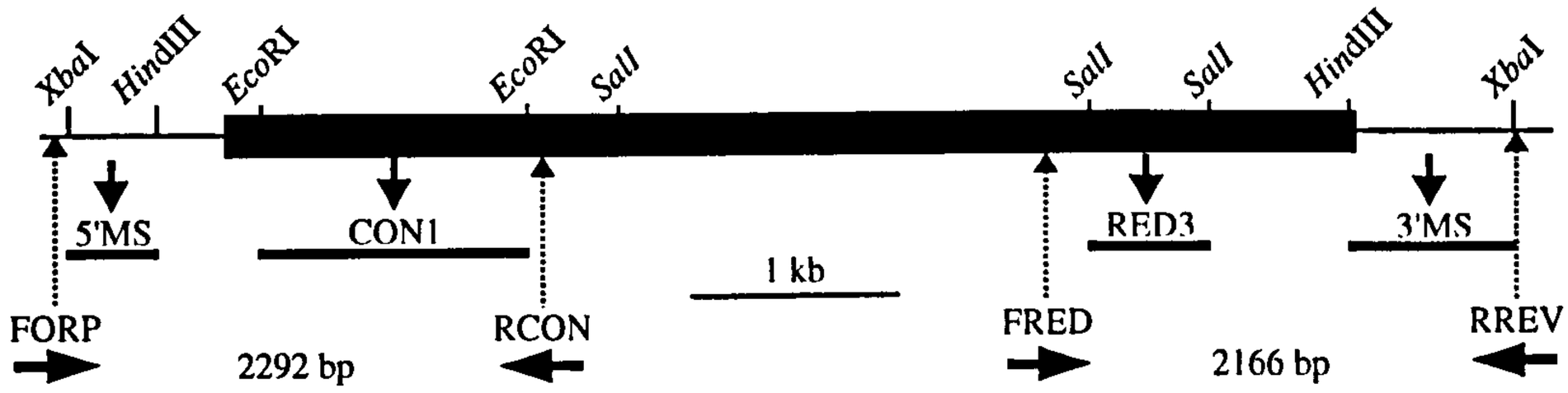


Figure 3.2. *MSAS* probes and PCR primers used in chromosome walking.

The black bar represents the *MSAS* coding sequence, including the single intron. 5'MS, CON1, RED3 and 3'MS are restriction fragments used as probes. Binding sites of PCR primers FORP (TTGGCTAGCTGTTCGAAGTC), RCON (ATTACCTAGTGGGTCTGGCT), FRED (CAATGGCTCACTGCCTATCA) and RREV (TACGCGTGGTATCATTCACC) are indicated by dotted arrows. The sizes of predicted PCR products are indicated between the primers.

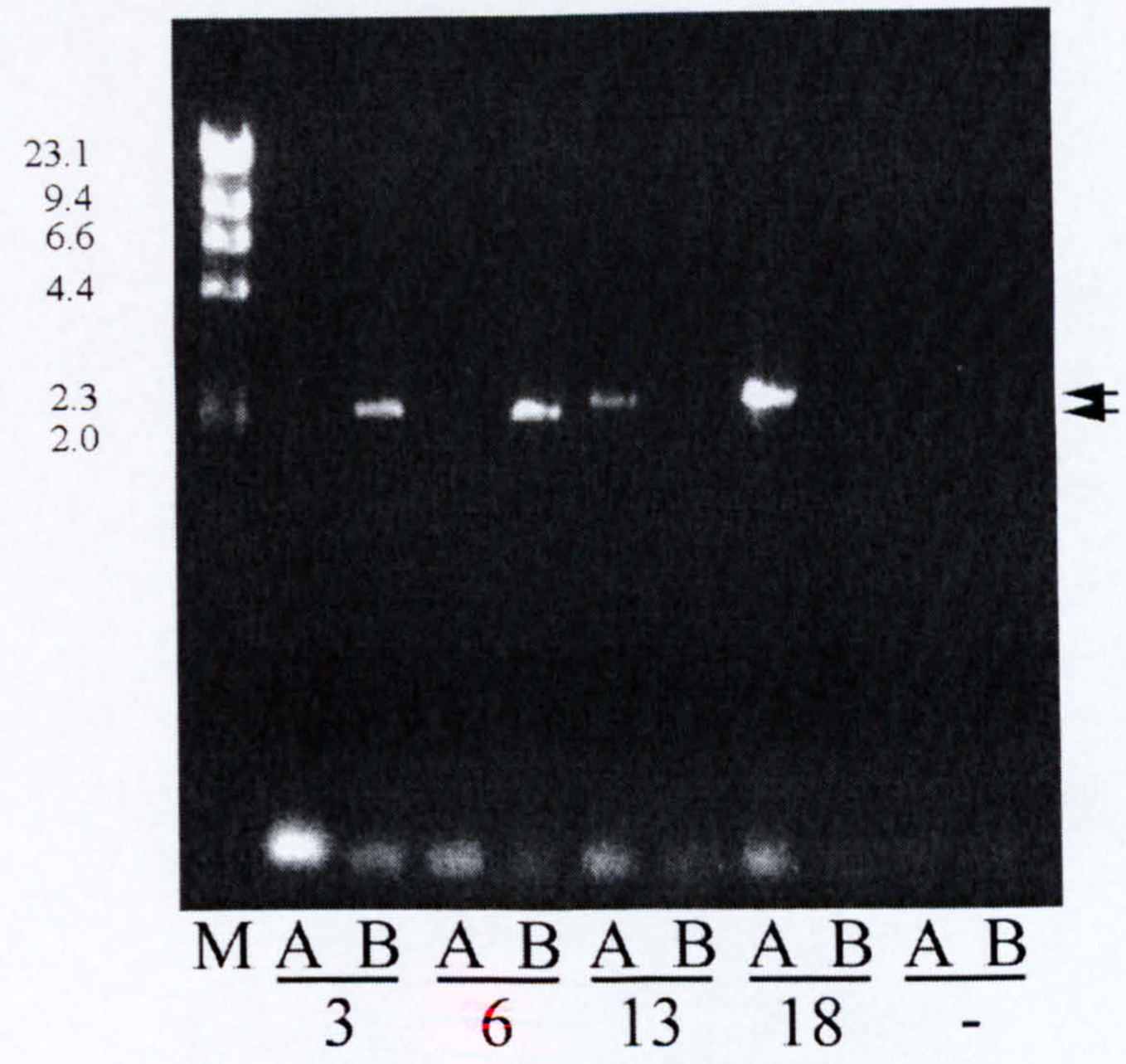


Figure 3.3. Screening phage clones using the polymerase chain reaction.

Overlapping of phage clones 3, 6, 13 and 18 with *MSAS* was assessed using two pairs of PCR primers on the phage template: **A**, FORP and CONREV; **B**, FRED and RREV. PCR products of 2.0 - 2.3 kb (arrowed) were obtained using primer pair A on clones 13 and 18 and using primer pair B on clones 3 and 6 in standard PCR reactions. Primers were annealed at 60° C. PCR products were fractionated on a 1% TAE gel at 70 mA. **M**, λ-*Hin*dIII DNA size markers (sizes of visible fragments indicated on left).

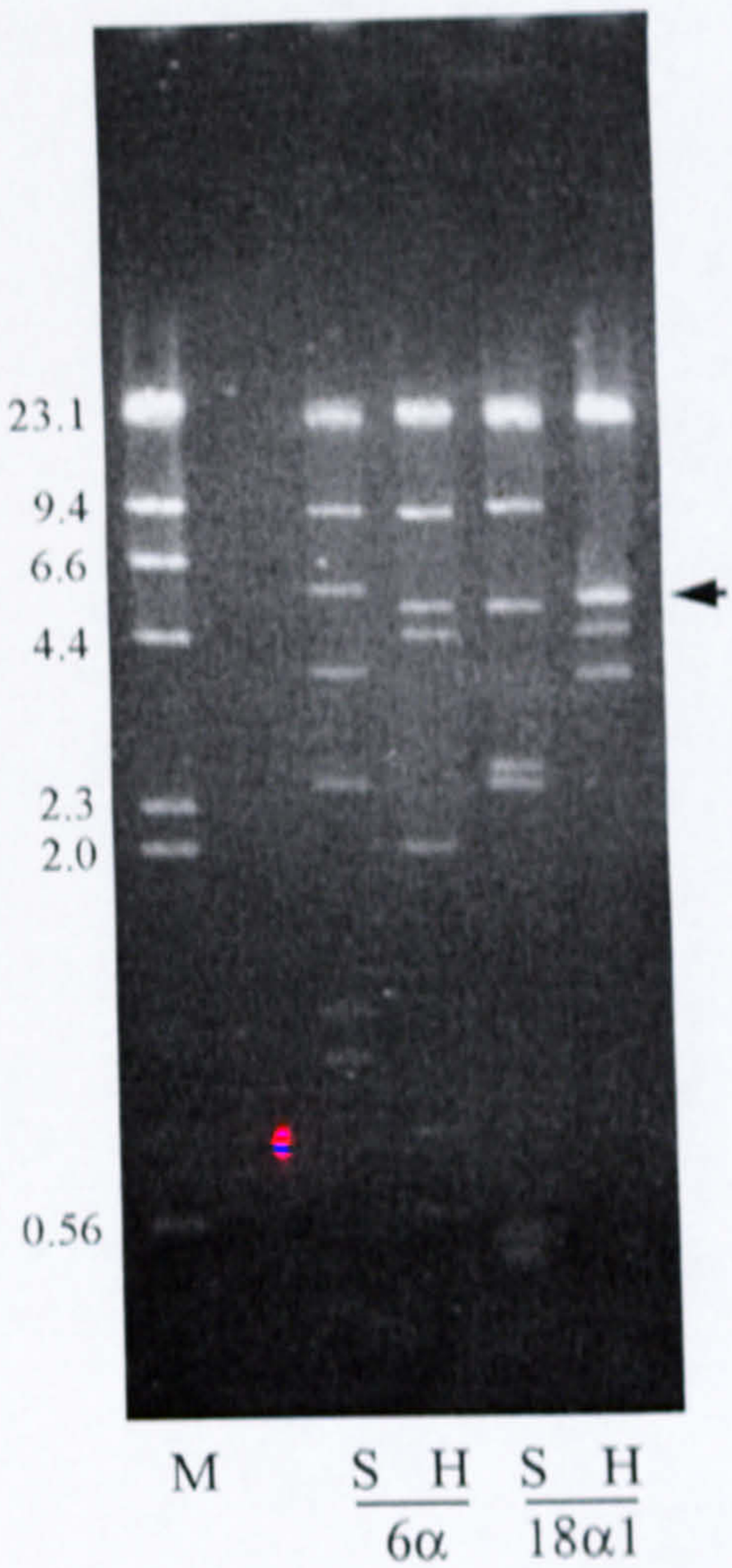


Figure 3.4 Restriction digests of 6α and 18α1 phage DNA.

Approximately 100 ng of phage DNA was digested with 10 U restriction enzyme in a total volume of 10 µl for 2 h at 37° C. The digestion products were fractionated on a 1% agarose gel in TAE buffer at 65 mA. **S**, *Sal*I digest; **H**, *Hin*dIII digest, **M**, λ-*Hin*dIII DNA size markers (sizes of fragments indicated on left). Restriction fragment sizes are listed in table 3.3. the 18α1/*Hin*dIII band arrowed consists of two co-migrating restriction fragments.

6α		18α1	
SalI	**HindIII**	**SalI**	**HindIII**
21	21	21	21
9.2	8.5	9.2	**5.2**
5.6	5.1	5.1	4.4
3.8	4.4	2.6	3.9
2.45	1.97	2.45	
1.15	0.77	0.54	
0.97	0.58	0.49	
0.54			

Table 3.3. Sizes in kb of restriction fragments produced by *SalI* and *Hin*dIII digests of phage clones 18α1 and 6α.

The figure highlighted in bold type represents two restriction fragments of similar size which comigrate on the agarose gel.

Probe	6α		18α1	
	SalI	**HindIII**	**SalI**	**HindIII**
5'MS	-	-	2.6	3.9
CON1	1.15	8.5	2.6	5.2
RED3	0.54	8.5	0.54	5.2
3'MS	3.8	5.1	-	-
SP6-20	1.15	8.5	0.49	5.2
T7-24	5.6	21	5.1	21

Table 3.4. 18α1 and 16α restriction fragments hybridising to *MSAS* probes and λGEM-11 arm-specific oligonucleotides.

were used to construct the restriction maps for the phage clones 18α1 and 6α shown in figure 3.5.

A further round of chromosome walking was undertaken to obtain clones overlapping the 5' end of 18α1 and the 3' end of 6α. Each clone was digested with *SalI* and digoxygenin-labelled RNA probes were transcribed from the λGEM-11 T7 RNA polymerase promoters. These riboprobes, in conjunction with 5'MS and 3'MS restriction fragments, were used to screen the *P. patulum* genomic library. Plaques hybridising to the T7/6α riboprobe but not to 3'MS, or to T7/18α1 but not to 5'MS were picked and purified by screening with the T7 riboprobes to give clones T7/18α1 and T7/6α. Time constraints prevented the characterisation of these fragments.

3.2.3 Transcript analysis

The two *SalI* restriction fragments extending 5' of *MSAS* in 18α1 and the two *SalI* fragments extending in the 3' direction from *MSAS* in 6α (figure 3.5) were obtained by fractionation of a restriction digest on an agarose gel followed by excision of the appropriate bands and elution of DNA from the gel slices. These restriction fragments were used to probe northern blots of total RNA extracted from *P. patulum* mycelium and conidiating cultures known to be expressing *MSAS* (figure 3.6). 18iii hybridised to transcripts of 6, 2 and 1.5 kb; 18iv hybridised to transcripts of 6 and 1.5 kb. This pattern of hybridisation indicates that at least two mRNAs are transcribed concurrently with *MSAS* from the 6 kb region of chromosome upstream of *MSAS*. The shorter (1.5 kb) transcript appears to arise from a coding region that is bisected by the *SalI* site immediately 5' of *MSAS*, and the longer (2 kb) transcript emanates from a coding region upstream of this restriction site. Fragments 6iii and 6iv hybridised only to the 6 kb *MSAS* transcript, indicating that the 8 kb region of chromosome downstream from *MSAS* contains no genes that are transcribed concurrently with *MSAS*. The hybridisation of 18iii and 6iii to the 6 kb transcript, presumably corresponding to *MSAS*, was probably due to the incomplete purification of the restriction fragments, i.e. their contamination by fragments 18iv and 6iv.

3.2.4 DNA sequence analysis of the chromosomal region upstream of *MSAS*.

Figure 3.5 shows the 4 kb *Hin*dIII restriction fragment and the 2 kb *Hin*dIII - *Sac*I restriction fragment upstream of *MSAS* that were subcloned from 18α1 into the pUBS1 and

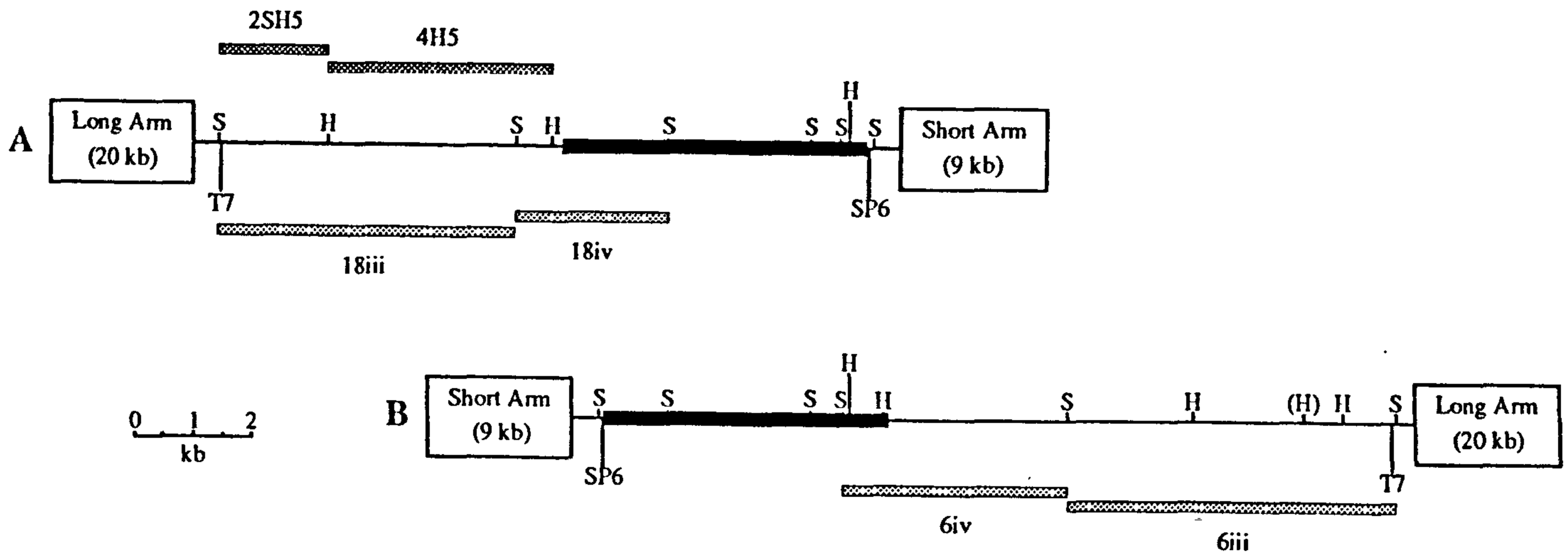


Figure 3.5 Restriction maps of λGEM-11 clones overlapping *MSAS*.

A. 18α1 (5' clone).

B. 6α (3' clone).

S, *Sal*I; H, *Hin*dIII; T7 and SP6, RNA polymerase promoters. The black bar represents the *MSAS* coding region and intron. Light shaded bars indicate *Sal*I restriction fragments used to probe northern blots; dark shaded bars indicate a 4 kb *Hin*dIII restriction fragment cloned into the pUBS1 plasmid vector to create p4H5 and the 2 kb *Hin*dIII - *Sac*I fragment cloned into pNEB193 to create p2SH5.

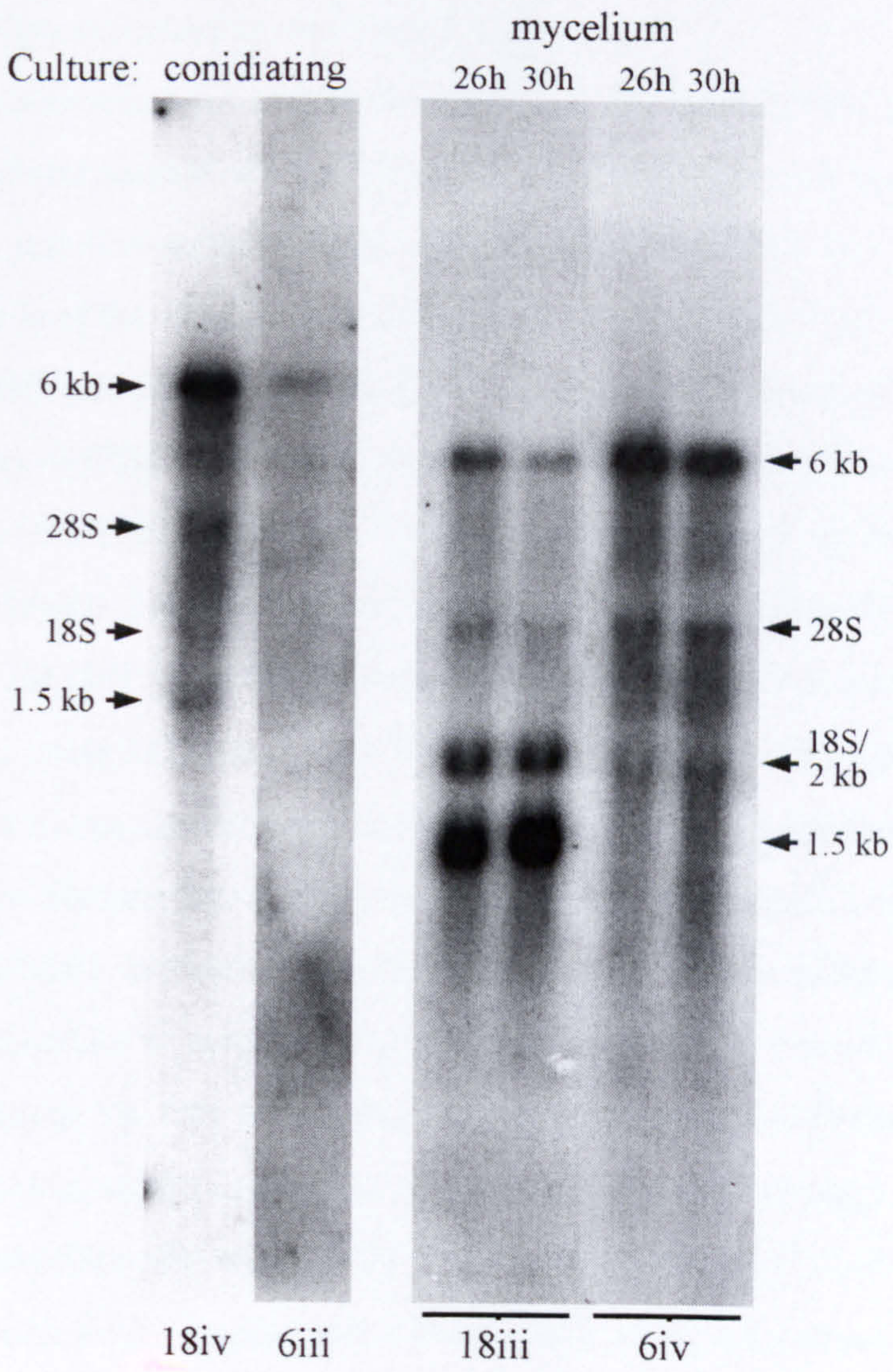


Figure 3.6 Northern blots of *Penicillium patulum* total RNA probed with restriction fragments from the phage clones 6α and 18α1.

Total RNA was extracted from 21 h conidiating culture (**18iv/6iii**), 26 h and 30 h mycelium (**18iii/6iv**), fractionated on formaldehyde denaturing agarose gels and blotted onto nylon membranes. These northern blots were probed with ^{32}P-labelled restriction fragments from phage clones overlapping *MSAS* (figure 3.5). The band at 6 kb represents hybridisation to the *MSAS* transcript. Bands labelled 28S and 18S probably represent degraded *MSAS* transcript co-migrating with 28S and 18S ribosomal RNA. In addition to these bands, fragment 18iv hybridises to a transcript of approximately 1.5 kb and fragment 18iii hybridises to two transcripts of approximately 2 kb (overlying the 18S band) and 1.5 kb.

pNEB193 vectors respectively, creating plasmids p4H5 and p2SH5. The identity of the insert in p4H5 was confirmed by using the plasmid to probe northern blots of *P. patulum* mycelium and conidiating cultures (figure 3.7). The plasmid hybridised to both the 2 kb and 1.5 kb transcripts described in section 3.2.3.

DNA sequence was obtained for both inserts (figure 3.23) and continuity between the two sequences on the chromosome was confirmed by sequencing of a 574 bp PCR product amplified from *P. patulum* genomic DNA using primers 2SHJUNT and 4H5JUNT, which anneal on either side of the *Hin*dIII site dividing the two fragments (figure 3.15).

Analysis of the p4H5 and p2SH5 DNA sequences indicated the presence of three large open reading frames (ORFs) on the opposite strand to *MSAS*. The 995 bp ORF nearest to *MSAS* was termed ORF1. The other two ORFs of 1001 bp and 455 bp were closely adjacent and, for reasons described below, were collectively designated ORF2. Predicted protein sequences for ORF1 and ORF2 were compared against the SwissProt protein sequence database, using MPsrch and FASTA programs. The first 60 amino acids encoded by ORF1 showed a strong similarity to the bovine and murine acetylcholine receptor ß subunit cytoplasmic domain between the transmembrane segments M3 and M4 (Tanabe *et al.*, 1984). The predicted products of the two open reading frames designated ORF2 showed strong similarities to neighbouring regions of proteins in the cytochrome P450 superfamily, indicating that they were part of a single gene. As cytochrome P450 monooxygenases are thought to catalyse a number of hydroxylation steps in the patulin biosynthetic pathway (Murphy *et al.*, 1974; Murphy & Lynen, 1975), ORF2 seemed a good candidate for a patulin pathway gene and subsequent research concentrated on characterising this gene.

3.2.5 Correlation of ORF1 and ORF2 to observed transcripts

PCR primers binding within ORF1 and ORF2 were used to amplify fragments from these ORFs for use as probes specific to these open reading frames (figure 3.8). To confirm that ORF1 and ORF2 corresponded with the two transcripts detected by northern blotting and also to determine the point of maximal expression of these two transcripts these probes were hybridised to a northern blot of total RNAs extracted from *P. patulum* mycelium of various ages (figure 3.9). All three transcripts were found to be strongly expressed between 28 h and 70.5 h.

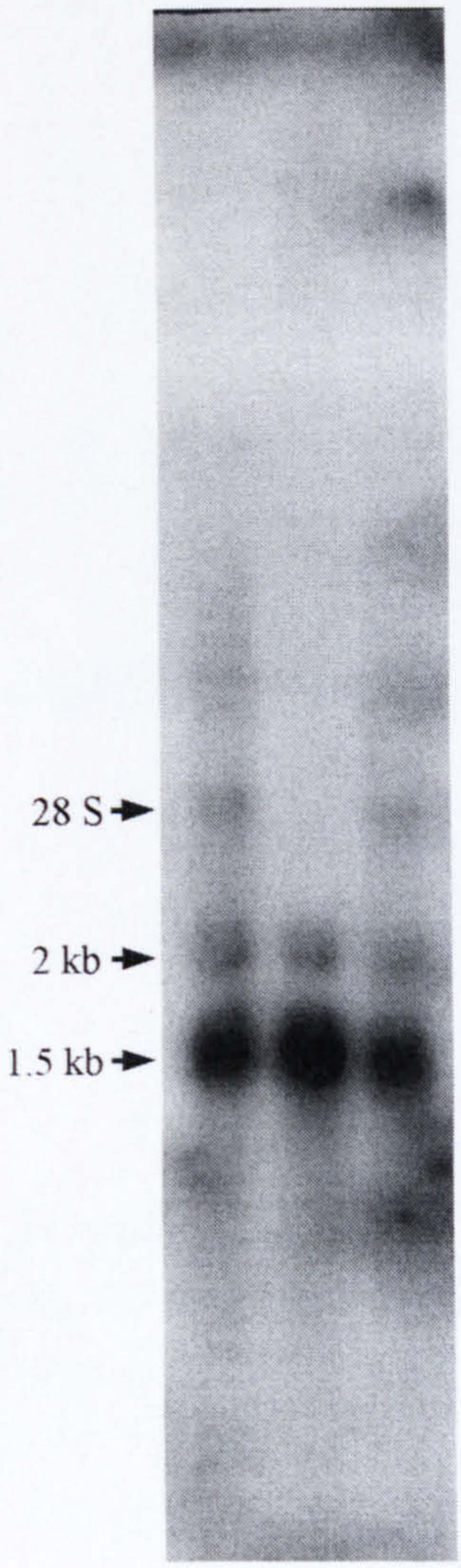


Figure 3.7 Northern blot of *Penicillium patulum* total RNA probed with p4H5.

Total RNA from 21 h conidiating cultures, 26 h and 30 h mycelium was probed with the ^{32}P-labelled insert of p4H5. The plasmid hybridised to two transcripts, of approximately 2 and 1.5 kb. The bands labelled 28S probably represent *ORF1/2* transcript co-migrating with 28S ribosomal RNA.

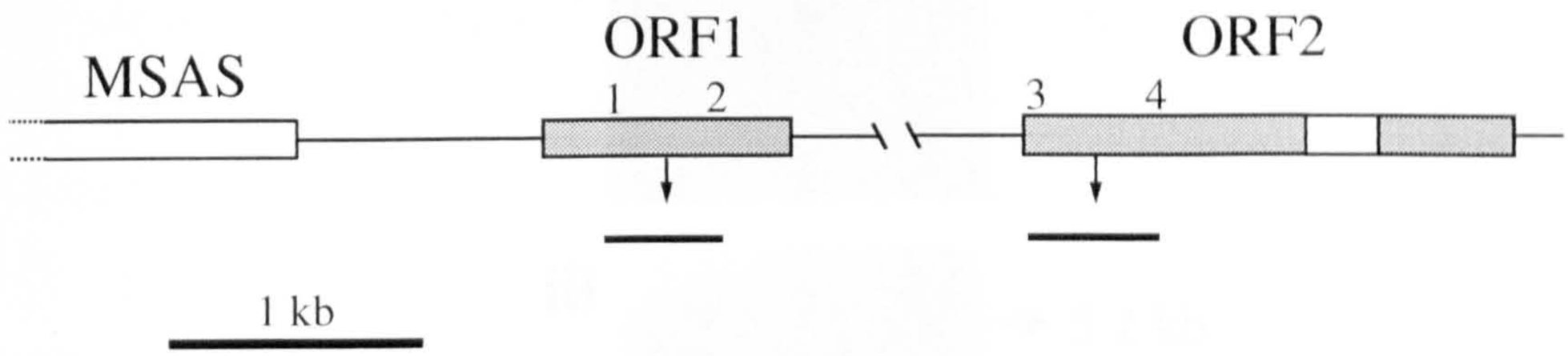


Figure 3.8 Derivation of DNA probes for ORF1 and ORF2 by PCR.

ORF1 and ORF2 are represented by grey boxes, drawn approximately to scale. The positions of the binding sites for primers 4H5F4 (CTTGAGTGCATCTGATT), 4H5R5 (TCTTGGCAACGGCAAAG), 4H5T3R2 (AAGACTTCATATCTAAC) and 4H5T3F1 (ATCGCTCACCATATTGC) are indicated by the numbers **1** to **4** respectively. PCR products are indicated by black bars. Standard PCR conditions were used (55° C annealing) and products were purified by elution from an agarose gel.

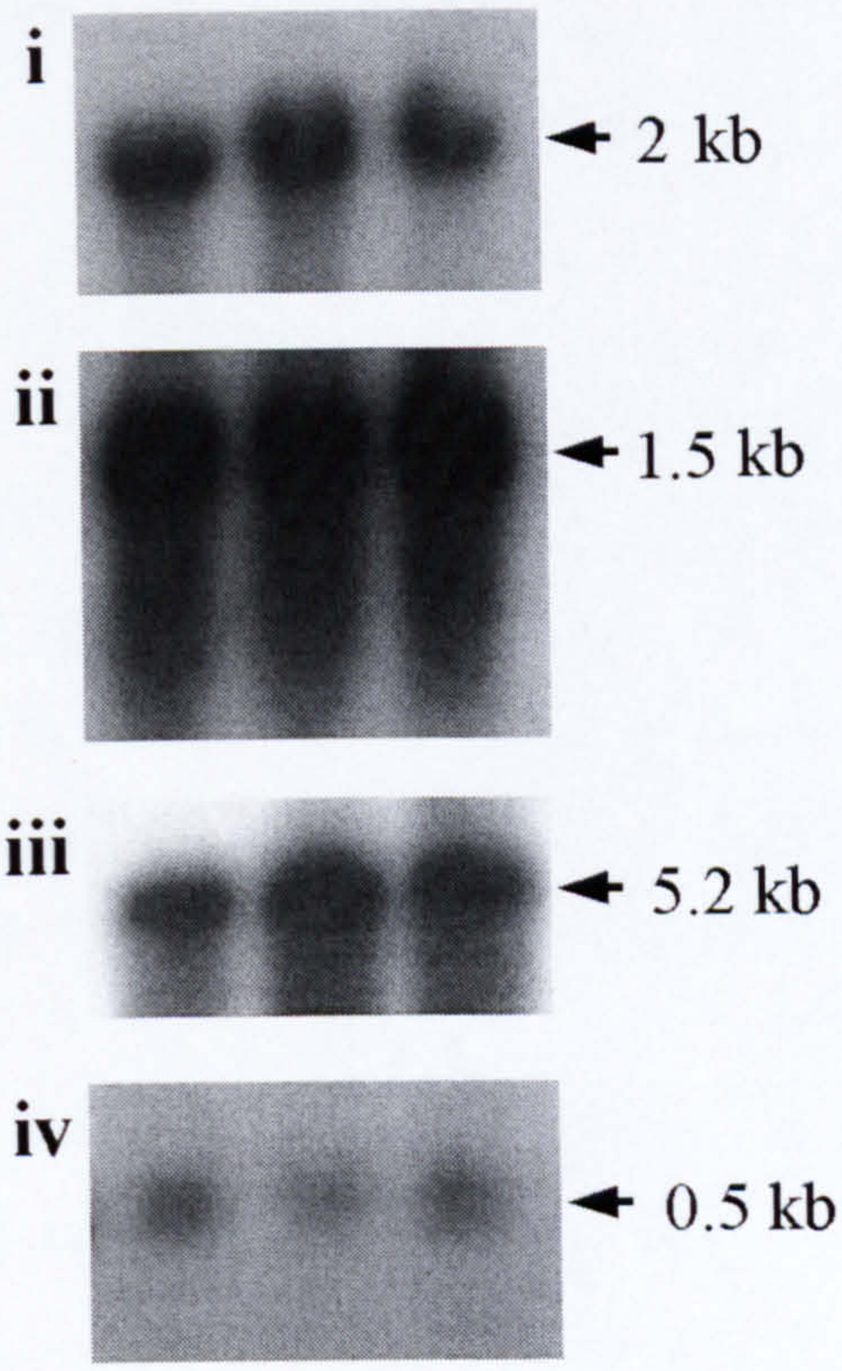


Figure 3.9 Northern blot of *Penicillium patulum* total RNA probed with *MSAS* restriction fragment, *ORF1* and *ORF2* PCR products.

Total RNA from 28 h (**1**), 45 h (**2**) and 70.5 h (**3**) *P. patulum* mycelium was probed with ^{32}P-labelled: **i.** ORF2 PCR product, **ii.** ORF1 PCR product, **iii.** *MSAS* condensing domain *Eco*RI fragment CON1 (figure 3.2) and **iv.** cytochrome c PCR product (control probe). Hybridisations indicate that transcripts corresponding to *MSAS*, *ORF1* and *ORF2* were present in all mycelium ages sampled.

The cytochrome c gene of *P. patulum* encodes a highly conserved respiratory electron carrier protein (Raitt *et al.*, 1994). Degenerate PCR primers CytC1 and CytC2 (figure 3.10, supplied by Dr A. Bailey) were used to amplify part of this gene for use as a control probe to assess the equivalence of total RNA loading in each lane of the blotted gel (figure 3.9). A product of the expected size was obtained in PCR reactions with a genomic template (figure 3.11) and gel-purified.

3.2.6 Characterisation of ORF2 cDNA

Poly(A)$^+$ RNA was prepared from 28 hour old *P. patulum* mycelium and used as a template for first strand cDNA synthesis. This oligo d(T)-primed cDNA was then used as the template in a PCR reaction with the primer 2SHF1 (figure 3.15) and the *Not*I-d$(T)_{18}$ primer (AACTGGAAGAATTCGCGGCCGCAGGAA$(T)_{18}$) that had primed cDNA synthesis; these amplified a major product of 1250-1300 b (figure 3.12). The band containing this product was eluted from the gel and used as the template in nested PCR reactions designed to confirm its identity as the ORF2 transcript and to provide some information about exon/intron structure. The primer combinations 2SHF1 with 2SHR3 and 2SHF1 with 2SHR4 both yielded amplified products from the RT-PCR-derived template that were 170 bp smaller than the products obtained from a genomic template (figure 3.13), implying the presence of one or more introns between the 2SHF1 and 2SHR3 primer binding sites. Comparison of the predicted protein sequence for *ORF2* to other cytochrome P450 proteins suggested that *ORF2* did not begin with the ATG codon at the 5' end of the long open reading frame (designated A), but somewhere further upstream. Two possible translation start sites upstream of this open reading frame were identified by homology to other cytochrome P450 proteins at 2831 bp (C) and 2901 bp (B). To ascertain whether each of these ATG codons lay within the transcribed region, PCR was carried out on a cDNA template using 5' primers designed to bind at each of the three possible start sites (figure 3.14); the primers BAM5'ORFA, BAM5'ORFB and 5'ORFC were each used in combination with 4H5T3F1 (figure 3.15). The sizes of the products obtained using these three primer combinations to amplify from both genomic and cDNA templates are shown in table 3.5. BAM5'ORFA amplified a cDNA product 50 bp smaller than the genomic product, BAM5'ORFB amplified a cDNA product 100 bp smaller than the genomic product and 5'ORFC did not amplify a major cDNA product but produced three minor products. These results show that the methionine codons at 2901 bp and 3099 bp were

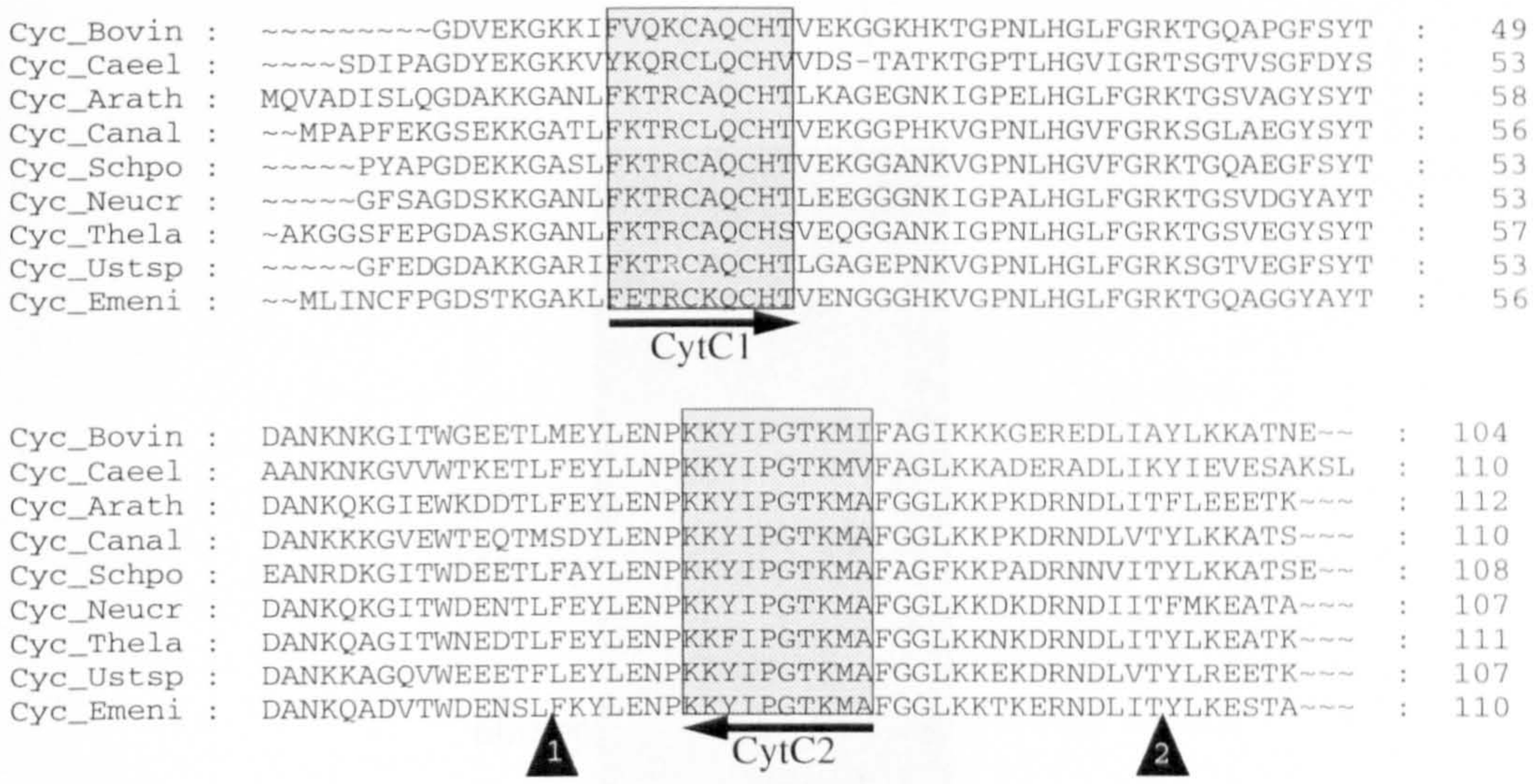


Figure 3.10 Conserved binding sites of cytochrome c degenerate PCR primers.

Protein sequence alignments created using the PILEUP program (Genetics Computer Group). Cytochrome c sequence identity codes: Bovin, bovine/pig/sheep; Caeel, *Caenorhabditis elegans*; Arath, *Arabidopsis thaliana*; Canal, *Candida albicans,* Schpo, *Schizosaccharomyces pombe*; Neucr, *Neurospora crassa*; Thela, *Thermomyces lanuginosa*; Ustsp, *Ustilago sphaerogena*; Emeni, *Emericella nidulans* (*Aspergillus nidulans*). Boxes indicate the regions encoded by CytC1 (TTYRARACNMGNTGYRMNGARTGYCAYAC) and CytC2 (GCCATYTTNGTNCCNGGDATRTAYTTYTT) primer binding sites (A. Bailey, pers. comm.). Arrows indicate orientation of primer binding. Black triangles indicate positions where the coding sequence is interrupted by small introns in *A. nidulans* (Raitt *et al.*, 1994).

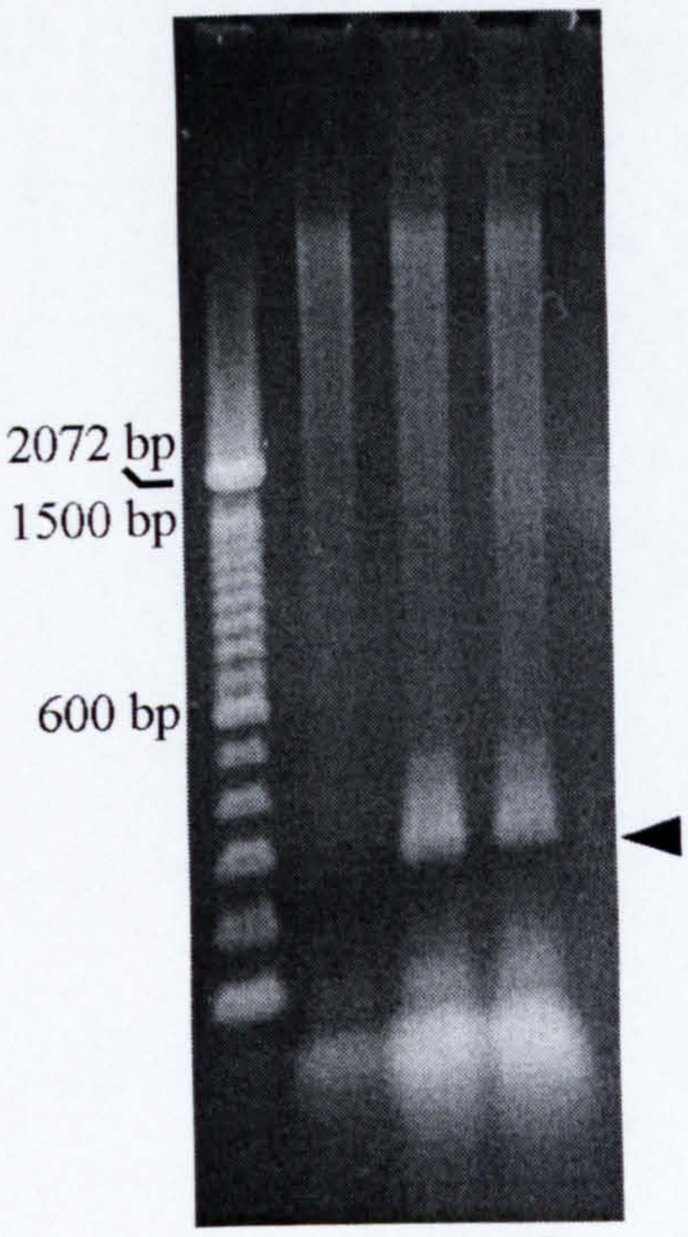


Figure 3.11 Amplification from the *Penicillium patulum* genome using cytochrome c degenerate PCR primers.

Primers CytC1 and CytC2 were used in PCR at three different concentrations: **A**, both at 0.4 μM; **B**, both at 1.2 μM; **C**, C1 at 3.16 μM, C2 at 2.36 μM. Primer annealing temperature was 55 °C. PCR products were fractionated on a 1.1% agarose gel in TAE buffer at 60-70 mA. **M**, 100 bp ladder DNA size marker (Gibco-BRL), sizes of brighter bands are indicated. An arrow indicates the product of 250-300 bp visible in lanes B and C.

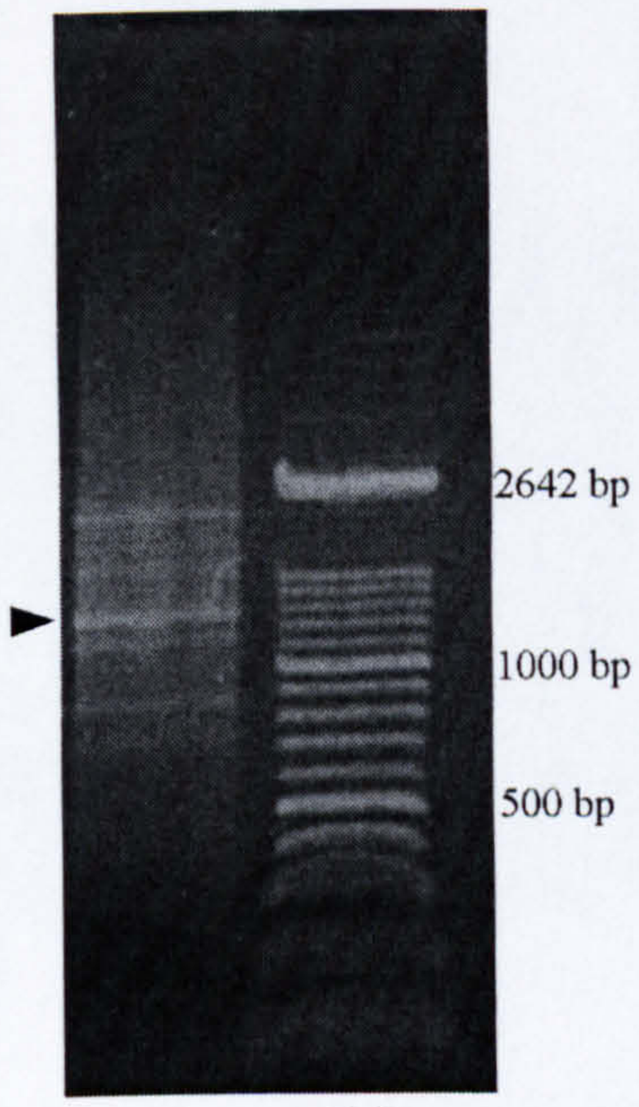


Figure 3.12 First-strand cDNA synthesis and PCR amplification from *Penicillium patulum* poly(A)$^+$ RNA.

Approximately 100 ng poly (A)$^+$ RNA was used as a template for first strand cDNA synthesis using the Ready-To-Go T-Primed first strand cDNA reaction mix (Pharmacia). Subsequently the complete cDNA reaction mix was used directly in PCR. Primers for this PCR were the oligo d(T) primer used in cDNA synthesis, and the *ORF2*-specific primer 2SHF1 (figure 3.15) which was added to a concentration of 0.4 µM. 2.5 U of Supertaq DNA polymerase (HT Biotechnology) was used in a total volume of 100 µl as described in the protocol supplied with the cDNA synthesis kit (Pharmacia). The thermal cycling program used for amplification was: (94° C, 2') x 1, (94° C, 30''; 55° C, 30''; 72° C, 1'30'') x 50, (72° C, 6') x 1. After amplification, a 10 µl aliquot of reaction mixture was fractionated on a 1.1 % agarose gel in 0.5 x TBE buffer. The major product (arrowed) was approximately 1250-1300 bp. **M**, 100 bp ladder DNA size marker (Boehringer-Mannheim), sizes of brighter bands are indicated.

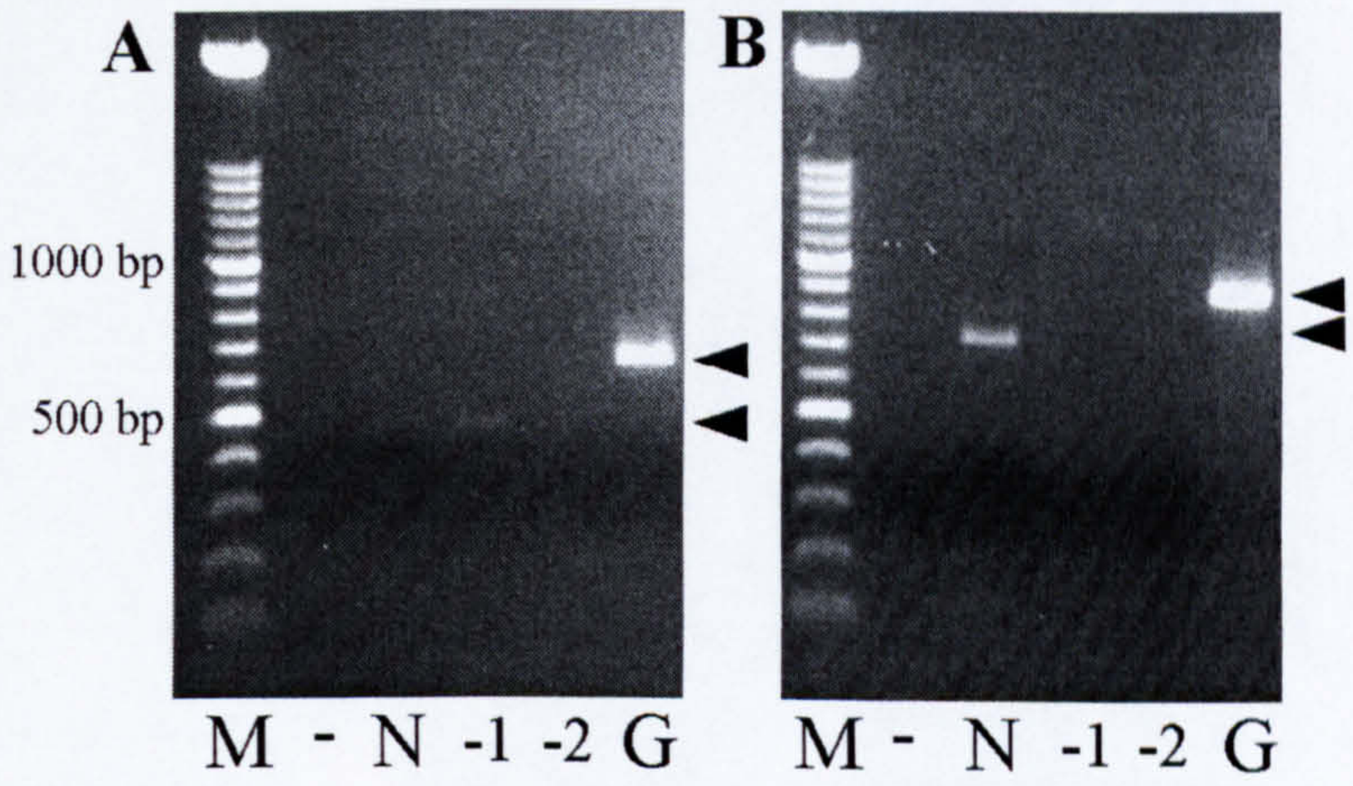


Figure 3.13 PCR amplifications within *ORF2* using *Penicillium patulum* genomic and amplified cDNA templates.

PCR using primers **A.** 2SHF1 and 2SHR3, **B.** 2SHF1 and 2SHR4 (figure 3.15). Standard *Taq* PCR conditions were used, with a primer annealing temperature of 55° C. The amplified cDNA template used was the gel-purified 1250-1300 bp product amplified from first strand cDNA (figure 3.12). PCR products were fractionated on a 1.2 % agarose gel in 0.5 x TBE buffer. **M**, 100 bp ladder DNA size marker (Boehringer-Mannheim, sizes indicated in panel **A**); -, negative (water) control PCR reaction; **N**, undiluted cDNA template; **-1**, 10^{-1} dilution of the cDNA template; **-2**, 10^{-2} dilution of the cDNA template; **G**, genomic DNA template (100 ng). Positions of bands are indicated by arrowheads.

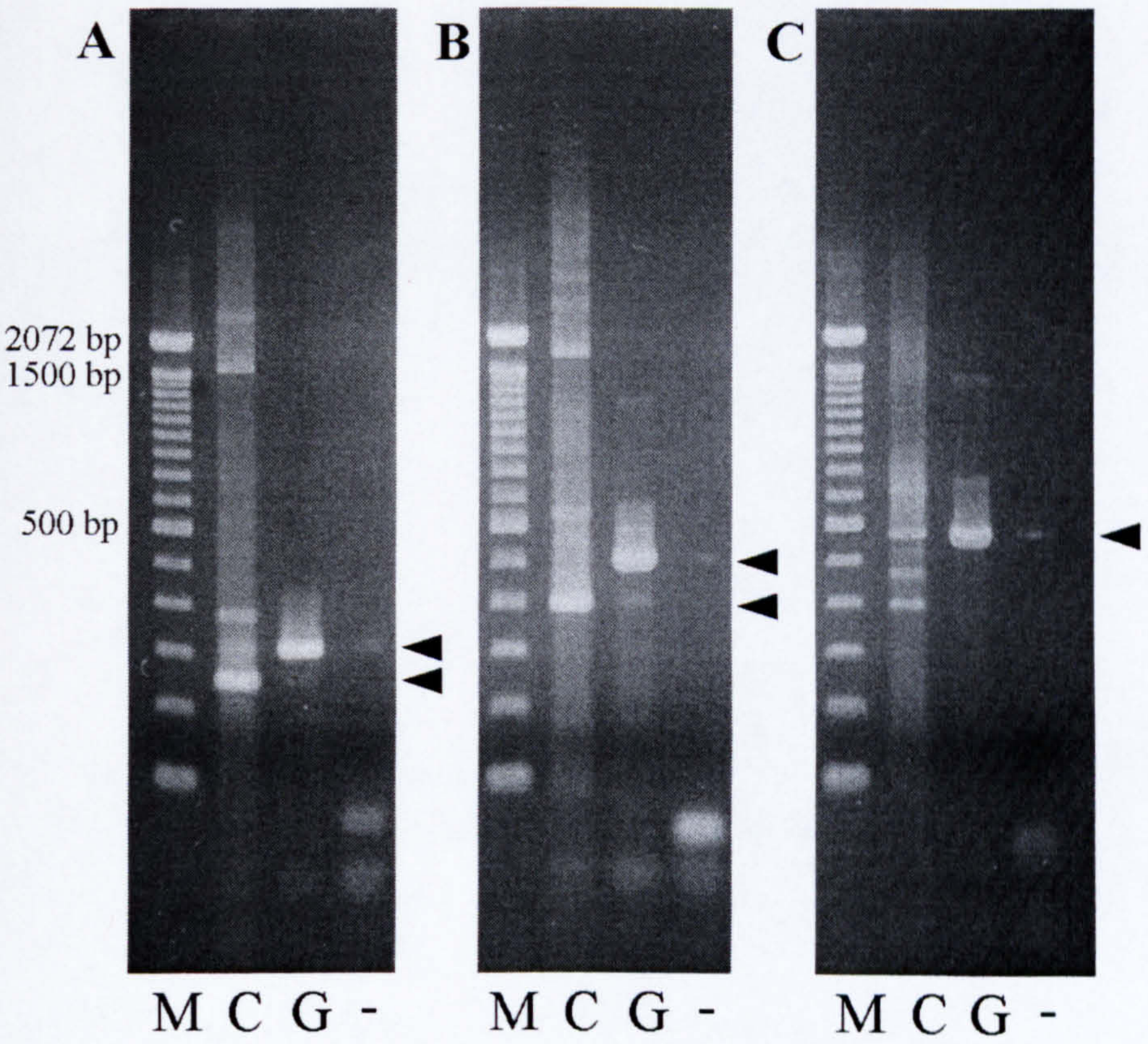


Figure 3.14 PCR using 5' primers binding at three potential translation start sites.

PCR reactions using three 5' primers (figure 3.15): **A.** BAM5'ORFA, **B.** BAM5'ORFB and **C.** 5'ORFC, in combination with the 3' primer 4H5T3F1. All primers were used at 0.4 µM, with dNTPs at 0.2 mM and 2.5 U Supertaq DNA polymerase (HT Biotechnology) in a 100 µl reaction. The thermal cycling program used for amplification was: (94° C, 5') x 1, (94° C, 1'; 55° C, 1'; 72° C, 1'30'') x 40, (72° C, 10') x 1. Aliquots (20 µl) of the PCR products were fractionated on a 1.5% agarose gel in 0.5 x TBE buffer at 70 mA. Major products are indicated by arrowheads. Templates were: **C**, 5 µl first strand cDNA; **G**, 100 ng genomic DNA; -, negative (water) control. **M**, 100 bp ladder DNA size marker (Gibco-BRL), sizes of brighter bands are indicated.

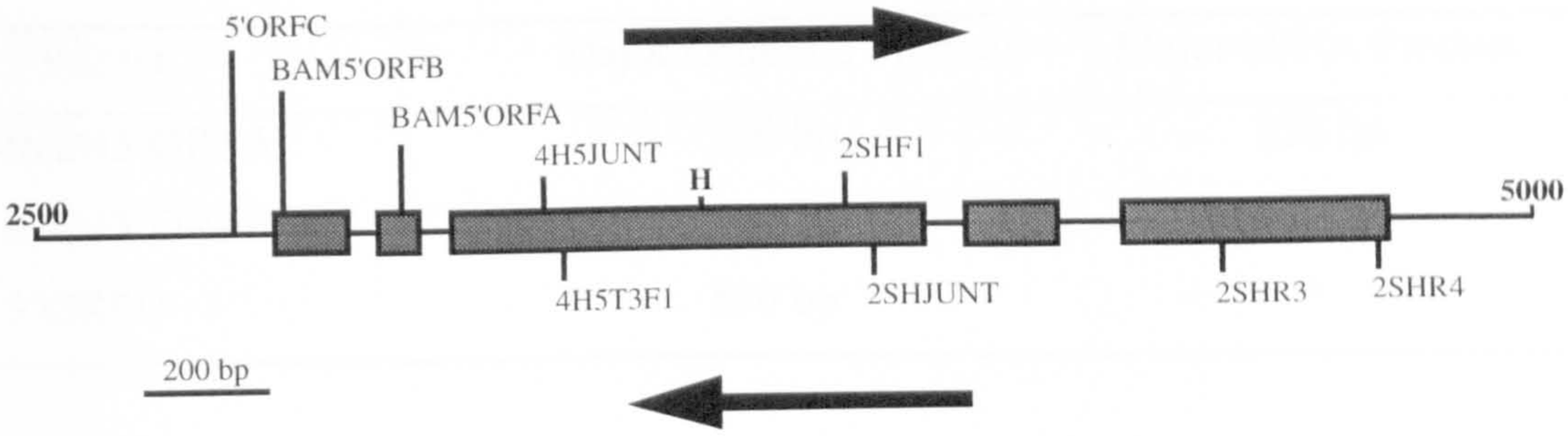


Figure 3.15 Positions of primer binding sites on *ORF2*.

The horizontal line represents the region of chromosome around *ORF2* (figure 3.23, 2500 bp-5000 bp) and the grey boxes represent the five exons of *ORF2*. H, *Hin*dIII restriction site.

The 5'-3' direction for primers labelled above and below the diagram is indicated by the arrows. Primer sequences and location on the genomic sequence shown in figure 3.23 are: BAM5'ORFA (CTGAGGATCCAATGAGCTGCAGGTGCCAAAAT, 3098-3119 bp); BAM5'ORFB (CTGCGGATCCTATGGATATCCTACAGTTCGC, 2899-2920 bp); 5'ORFC, (CGCGTTGATGAGAACAATAA, 2824-2843 bp); 4H5T3F1 (ATCGCTCACCATATTGC, 3396-3379 bp); 2SHF1 (ATTCAATGAAACAGACC, 3850-3866 bp); 2SHR3 (CCAATATGATCCCTCCG, 4491-4475 bp); 2SHR4 (GAGCGAACACCTCCTGT, 4748-4732 bp); 4H5JUNT (TATCGCACCTTATTACGCAG, 3340-3359 bp); 2SHJUNT (TTCCATCAGAACTCCTCCTA, 3914-3895 bp).

5'Primer	Major Genomic Product	Major cDNA Product
BAM5'ORFA	300 bp	250 bp
BAM5'ORFB	500 bp	400 bp
5'ORFC	550 bp	*

Table 3.5 PCR products obtained using 5' primers binding at possible translation initiation regions in combination with 4H5T3F1.

Primer 5'ORFC did not amplify a major product with a cDNA template (*).

within the poly(A)$^+$ transcript and suggest that the codon at 2831 bp is situated upstream of, or very close to, the transcriptional start point. The differences between genomic and cDNA product sizes indicated the presence of an intron of approximately 50 bp between primers BAM5'ORFA and 4H5T3F1 and another of the same length between primers BAM5'ORFA and BAM5'ORFB.

Based on the above results and comparison to other cytochrome P450 proteins, the methionine codon (B) at 2901 bp (figure 3.23) was assumed to be the start of translation. A putative full length cDNA (of approximately 1600 bp) was amplified by PCR using the Expand^{TM} high fidelity enzyme mixture with primers BAM5'ORFB and ECO3'ORF2 (see section 3.2.7). Partial DNA sequencing of this PCR product revealed the presence of four small introns in this gene, arranged as shown in figures 3.16 and 3.23. The predicted start of translation for *ORF2* would result in a coding sequence of 1581 bp encoding a protein product with a predicted molecular mass of 58.4 kDa.

3.2.7 Expression of full-length ORF2 cDNA as a glutathione S-transferase fusion protein.

Amplification of the putative full length copy of the ORF2 cDNA with the tailed primers BAM5'ORFB (CTGC**GGATCC**TATGGATATCCTACAGTTCGC) and ECO3'ORF2 (CGGC**GAATTC**TAATCAAATTGAGCGAACA) introduced a *Bam*HI site at the 5' end and an *Eco*RI site at the 3' end (restriction sites in the primer sequences are highlighted in bold type). The PCR product was digested with *Bam*HI and *Eco*RI and ligated into the glutathione S-transferase (GST) fusion expression vector pGEX-BHE(E)XSX (supplied by Dr S. Screen), a derivative of pGEX-4T-3 (Pharmacia) with a modified multiple cloning site (figure 3.17). The resulting construct, pGEX-ORF2, was used to transform *E. coli* DH5α.

Expression of pGEX-ORF2 in small (5 ml) cultures was investigated using five different recombinant *E. coli* DH5α clones, numbered 2, 3, 8, 12 and 13. Expression of the full-length ORF2 cDNA as a GST fusion protein of the predicted size (84.4 kDa) was observed in clones 3, 8, 12 and 13 by SDS-PAGE (figure 3.18). Each of these clones may contain errors in the nucleotide sequence introduced during the PCR, but any such mutations have not resulted in the interruption of the open reading frame by a stop codon. Clone 2 expressed a truncated fusion protein of between 30 and 46 kDa, presumably due to the

introduction of an alteration in the nucleotide sequence which changed an amino acid-encoding codon to a stop codon.

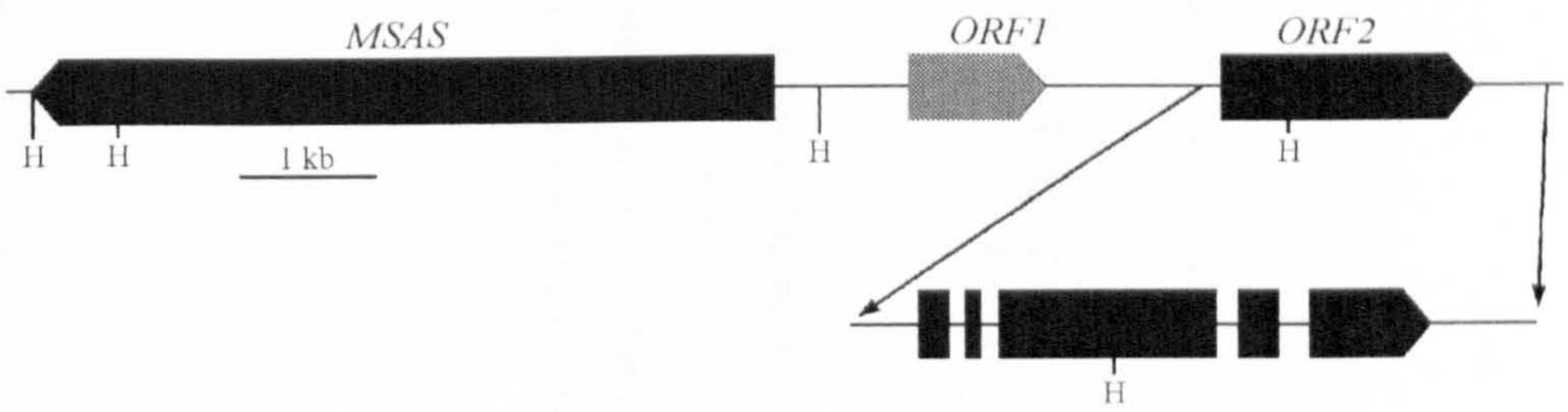


Figure 3.16 Two open reading frames clustered with *Penicillium patulum MSAS*.

The upper part of the diagram is a restriction map of the region of *P. patulum* chromosome containing *MSAS* and the two clustered open reading frames, designated *ORF1* and *ORF2*. Genes represented by black wedges (*MSAS* and *ORF2*) have defined intron/exon structures while the gene represented by the grey wedge (*ORF1*) has not been fully characterised. Arrowheads indicate direction of transcription. **H**, *Hin*dIII restriction site. The lower part of the diagram shows the exon structure of *ORF2*.

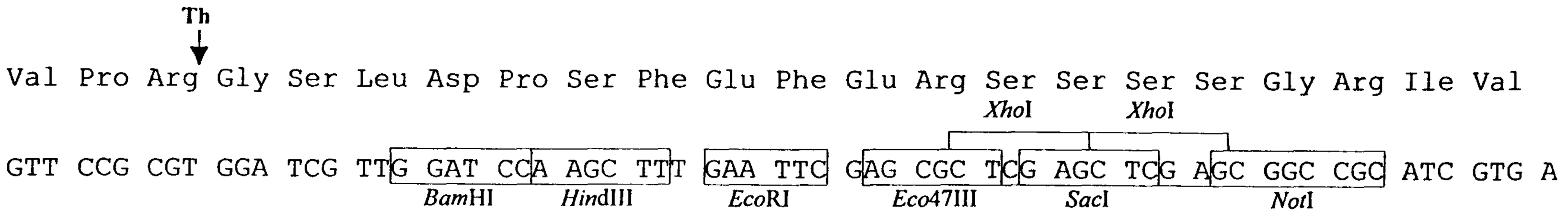


Figure 3.17 Sequence of the multiple cloning site for GST fusion expression vector pGEX-BHE(E)XSX.

This vector, supplied by Dr. S.E. Screen, is a derivative of pGEX-4T-3 (Promega) with a modified multiple cloning site as shown.

Amino acid sequence for the protein product of the parent vector is shown above the nucleotide sequence. Restriction sites and the thrombin cleavage site for the fusion protein (Th) are indicated above and below the sequence.

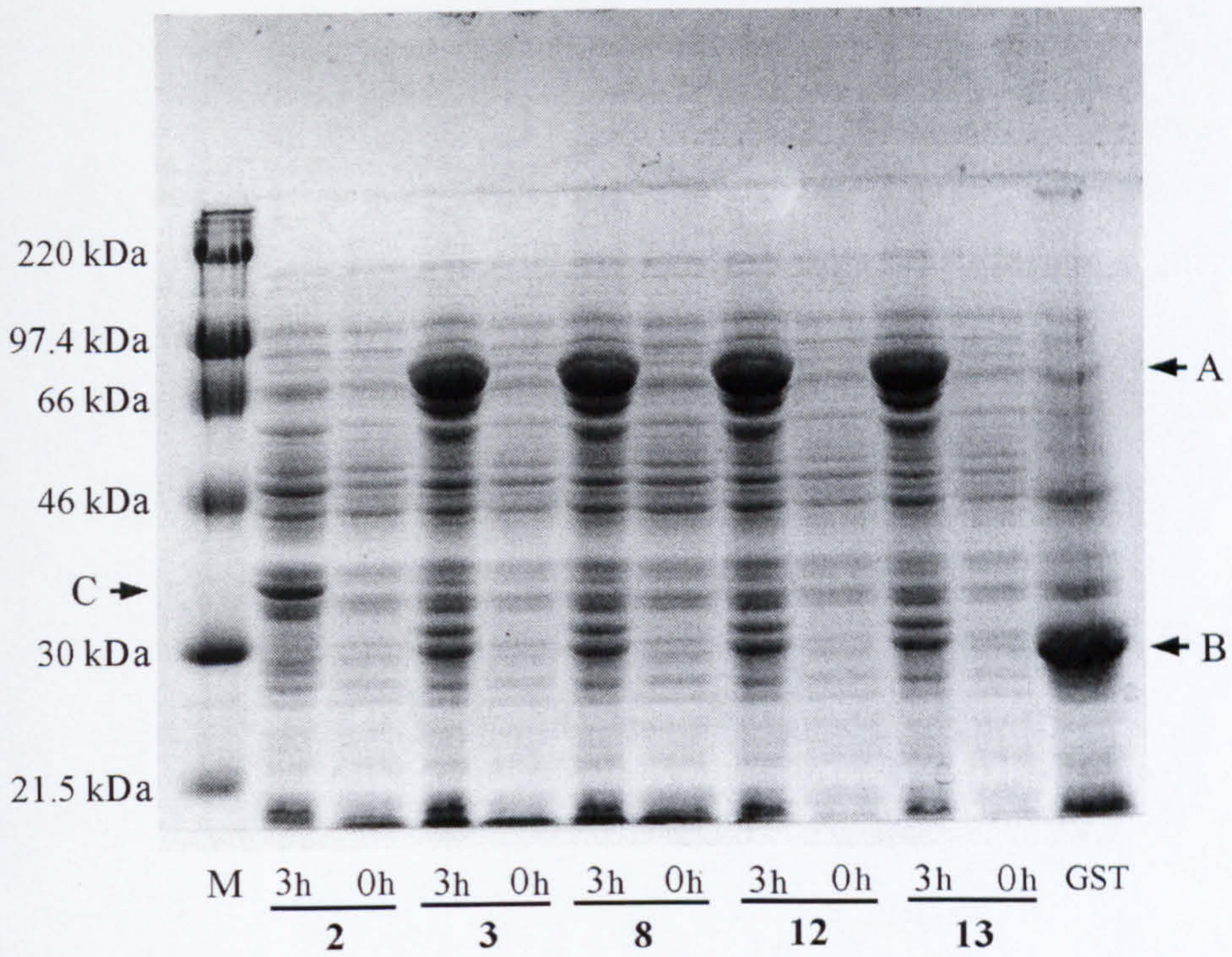


Figure 3.18 SDS-PAGE of products of expression of pGEX-ORF2 constructs in *E. coli.*

Samples were taken from each of 5 cultures (pGEX-ORF2-**2**/**3**/**8**/**12**/**13**) before induction (0h) and 3 h after induction with 1 mM IPTG. Clones **3**, **8**, **12** and **13** show expression of an 82 kDa GST-ORF2 fusion protein (**A**). **GST,** control sample expressed using non-recombinant pGEX vector showing GST protein of 26 kDa (**B**). Note the truncated fusion protein expressed by clone pGEX-ORF2-2 (**C**). **M**, Rainbow™ molecular weight markers (Amersham) with sizes of the bands indicated at the side.

3.3 DISCUSSION

3.3.1 Identification of linked genes expressed concurrently with *MSAS.*

The existence of at least two genes linked to *MSAS* has been demonstrated by the chromosome walking and northern blotting described in sections 3.2.2 and 3.2.3. These genes appeared to be transcribed at all the time points and developmental stages assayed; therefore expression of these genes cannot be said to be co-ordinated with *MSAS* expression and could even be constitutive. However, in the light of the reported clustering of secondary metabolic pathway genes in filamentous fungi, it seems likely that genes clustered with and expressed concurrently with *MSAS* are involved in the patulin biosynthetic pathway. Sequencing of the cloned region of chromosome upstream of *MSAS* revealed three long ORFs. The nearest of the three to *MSAS* was designated *ORF1,* while the remaining two adjacent ORFs showed strong homology to adjacent regions of enzymes of the cytochrome P450 type and were therefore collectively designated *ORF2.*

3.3.2 Structure of *ORF2.*

RT-PCR indicated that both the ORFs of *ORF2* were transcribed to a single mRNA and were interrupted by several introns (section 3.2.6). Sequencing of a cloned cDNA led to the identification of four small introns between the predicted start and end of translation (shown in figures 3.16 and 3.23). The majority of sequenced genes from filamentous fungi have been found to contain one or more introns, of average length 69 bp, and lengths between 50 bp and 250 bp have been commonly reported (Gurr *et al.*, 1987; Unkles, 1992). The four introns observed in *ORF2* are all within this size range. The 5' splice site consensus is GTRNGY (Unkles, 1992), where R is a purine, Y is a pyrimidine and N is any base, and this sequence is found in all *ORF2* introns except intron II (table 3.6). The 3' splice site consensus is YAG and this is followed in all *ORF2* introns (table 3.6). The consensus internal element for lariat formation in the filamentous fungi is NNCTRAY (Unkles, 1992) and this sequence, or a close approximation to it, can be observed in all ORF2 introns at 10-15 bp from the 3' end (table 3.6).

Intron	5' splice site	3' splice site	Putative internal consensus (bp from intron 3' end)
ORF2 I	GTATGT	CAG	ATCTAAC (10)
ORF2 II	GTAAAG	TAG	TGCTAAG (12)
ORF2 III	GTATGC	CAG	CGCTAAT (15)
ORF2 IV	GTCTGT	TAG	GGCTAAT (12)
MSAS	GTCTGT	TAG	TACTGAC (9)
FAS2 intron 1	GCAAGT	TAG	AGCTAAC (10)
FAS2 intron 2	GTAAGT	TAG	TGCTAAC (10)
NRFA	GTAAGA	TAG	AGCTGAT (18)

Table 3.6 Intron consensus elements in characterised *P. patulum* protein encoding genes.

Consensus elements for *ORF2* introns are shown with those from *MSAS* (Beck *et al.*, 1990), *FAS2* (Wiesner *et al.*, 1988) and *NRFA* (Ellis, 1996) for comparison.

3.3.3 Transcriptional control motifs in *ORF2.*

The promoters of fungal genes contain a number of regions involved in transcription control. These typically include upstream activation or repression sequences (UAS/URS) involved in the regulation of gene expression and a pyrimidine-rich region, known as the CT box, which is involved in initiation of transcription (Punt *et al.*, 1992; Unkles, 1992). CT boxes (pyrimidine-rich tracts of between 10 and 60 bp) are often situated immediately before the major start of transcription in the genes of filamentous fungi, with transcription initiation occurring at the first purine downstream (Fischer *et al.*, 1995; Punt *et al.*, 1992). Pyrimidine rich motifs of 11 and 17 bp in length can be identified at -102 and -72 bp from the putative start of translation (figure 3.23, 2789 bp and 2813 bp), however a transcriptional start point immediately downstream of either of these motifs would not be consistent with the suggested start of translation.

TATA box sequences, AT-rich motifs with the consensus TATAAA involved in binding of the TFIID transcription factor in many yeast and higher eukaryotic genes, can be identified in the promoters of most fungal genes between 40 and 100 bp from the transcriptional start point and between 50 and 150 bp from the start of translation (Unkles, 1992). Experiments involving deletion of these motifs in filamentous fungal promoters have not always supported a role for these sequences in initiation of transcription (Punt, 1992). Several sequences resembling the TATA consensus can be identified in the region 5' of *ORF2*, but the closest one to the putative site of translation initiation is TATAT at 188 bp upstream (figure 3.23, 2710 bp).

Nitrogen metabolite repression of a number of genes in filamentous fungi is known to be mediated by the binding of a nitrogen regulatory factor to specific DNA sequences upstream of the gene (Caddick *et al.*, 1994). Transcription of genes in the patulin biosynthetic pathway has been shown to be regulated by the action of such a protein (Ellis, 1996; GM Gaucher, personal communication). Other nitrogen regulatory factors studied in filamentous fungi, such as NIT2 in *Neurospora crassa*, NRE in *Penicillium chrysogenum* and AREA in *Aspergillus nidulans* are zinc finger proteins belonging to the GATA family of DNA-binding proteins, so called because they bind to the core DNA sequence GATA (Caddick *et al.*, 1994; Haas & Marzluf, 1995; Orkin, 1992). Experiments in *N. crassa* and *P. chrysogenum* have shown that the binding sites for such regulatory proteins must contain at least two GATA sequences, in any orientation, separated by 3-40 bp (Chiang & Marzluf, 1994, Haas & Marzluf, 1995). Twelve GATA core sequences can be identified in

the 1.2 kb region of chromosome 5' of ORF2 and there are three instances where two GATA sites are situated within 40 bp of each other (figure 3.23).

Biosynthesis of patulin is thought to be subject to manganese-regulated transcriptional control (Scott *et al.*, 1986; Scott, Jones & Gaucher, 1986; section 1.1.4). The most thoroughly characterised eukaryotic metal-regulated system is metallothionein biosynthesis (Thiele, 1992). Metallothionein promoters possess multiple copies of *cis*-acting metal regulatory elements (MREs), consisting of a series of 13-15 bp imperfect repeats, which are thought to be bound by a metal responsive transcription factor (Thiele, 1992). In the basidiomycete *Phanerochaete chrysosporium*, expression of the two isozymes of manganese peroxidase is known to be subject to manganese regulation at the level of gene transcription (Brown *et al.*, 1991) and the promoters of these genes (*mnp1* and *mnp2*) contain several putative MREs which fit the core consensus sequence (TGCRCNC) for the higher eukaryotic MRE (Mayfield *et al.*, 1994). Nine putative MREs can be identified within the 1.2 kb region of chromosome 5' of ORF2, all of which differ from the consensus sequence by a single nucleotide. In two cases, pairs of these putative MREs (2405-2414 bp and 2581-2590 bp, figure 3.23) overlap and form a 4 bp palindrome (TGCA) as has been observed in the *mnp1/2* promoters (Mayfield *et al.* 1994).

3.3.4 Polyadenylation of the *ORF2* transcript

In animals, the sequence AATAAA is a highly conserved polyadenylation signal motif, occurring 10-30 bp 5' of the polyadenylation site (Humphrey & Proudfoot, 1988). In plants, the polyadenylation signal motif is less conserved (AAUNNN) and multiple copies may be present (Gallie & Bailey-Serres, 1997). While this motif is present in the expected position in a number of fungal genes, in many it is either absent or differently situated relative to the polyadenylation site (Unkles, 1992). No such sequence is present in the vicinity of the polyadenylation site in the *P. patulum* genes *FAS2* and *MSAS* (Beck *et al.*, 1990). A perfect AATAAA motif is present in the putative 3' non-coding region of *ORF2*, sited 343 bp downstream of the stop codon. The poly(A) tail added to eukaryotic mRNA molecules is usually about 200 bp in length (Lewin, 1997) which would result in a predicted transcript length for *ORF2* of approximately 2200 bp, consistent with the transcript size observed in northern blots.

3.3.5 Initiation and termination of translation sites for *ORF2*.

The amino terminus of the eukaryotic microsomal cytochrome P450s serves as a membrane-insertion and halt-transfer signal. This will typically consist of an acidic residue closely following the initiation methionine followed by 14 to 20 hydrophobic residues and then several basic residues (Gonzalez, 1989). The methionine codon at 2901 bp was identified as the most likely start of translation as this would result in just such an amino terminus to the protein (figure 3.19); this prediction was supported by the results of RT-PCR at the 5' end of the *ORF2* transcript (section 3.2.6). Translation of most fungal genes begins at the first AUG codon of the mRNA (Gurr *et al.*, 1987), but the AUG will only be recognised efficiently as an initiation codon when it is in the context of an appropriate sequence environment, which in eukaryotes has the consensus GCCRCCAUGG. The purine three bases before the AUG and the G immediately following it are the most important bases in this consensus, influencing translational efficiency by a factor of ten, while the other bases have much smaller effects (Lewin, 1997). Filamentous fungi show a strong preference (64%) for A at position -3 (Gurr *et al.*, 1987; Unkles, 1992). The putative initiation codon for *ORF2* is within an approximation to the consensus ribosome binding site (AACACUAUGG).

A putative stop codon for *ORF2* has been identified which would result in a gene product of a similar size to other cytochrome P450 enzymes (figure 3.23, 4757-4759 bp). Eukaryotes exhibit a strong bias for a purine immediately following the stop codon (Cavener & Ray, 1991). The putative stop codon identified for *ORF2* has a pyrimidine (cytosine) in this position, found in about 30% of filamentous fungal genes (Unkles, 1992).

3.3.6 ORF2 is proposed to be a cytochrome P450 monooxygenase.

The predicted amino acid sequence encoded by *ORF2* was compared against protein sequences in the SwissProt database using both FASTA and Mpsrch programs to search for the best global alignments and local matches respectively. Both methods revealed a strong similarity between ORF2 and various plant *trans*-cinnamate 4-monooxygenases and other cytochrome P450s, with *Phaseolus aureus* CYP73 and Rat CYP2J3 showing the highest similarity scores. The accepted classification system for cytochrome P450s defines a member of one gene family as sharing an amino acid identity of at least 40% with members

Figure 3.19 Putative ORF2 amino terminal membrane insertion-halt transfer signal sequence. Sequence shows amino acids encoded by predicted first exon of *ORF2*. Residues in bold print are acidic, residues in light shaded boxes are hydrophobic and residues in darker shaded boxes are basic (Matthews & van Holde, 1990).

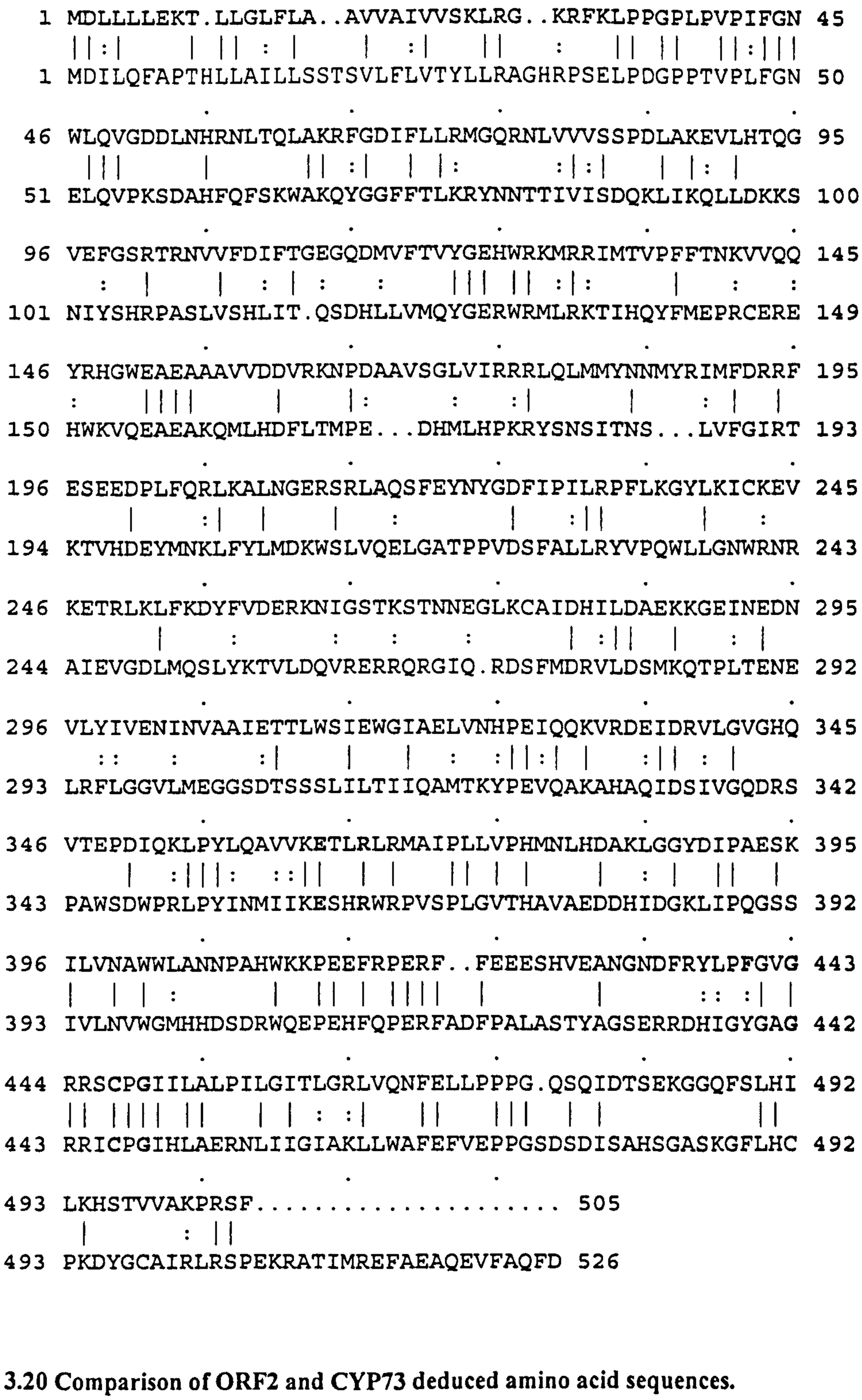


Figure 3.20 Comparison of ORF2 and CYP73 deduced amino acid sequences.

The upper sequence is CYP73 from *Phaseolus aureus* (Mizutani *et al.*, 1993), lower sequence is ORF2. Vertical lines indicate identity between amino acid residues, two dots indicate conserved amino acid similarities. Amino acid similarity between the two sequences is 38 %, amino acid identity is 27 %. Sequences were aligned using the GAP program (Genetics Computer Group). The heme binding motif FXXGXXXCXG and the EXXR motif conserved among cytochrome p450s (Nelson *et al.*, 1993; Yu *et al.*, 1997) are highlighted in bold type.

of the same gene family (Gonzalez, 1989; Nelson *et al.*, 1993). As CYP73 shares only 27% amino acid identity (figure 3.20), then the cytochrome P450 encoded by ORF2 is proposed to be a member of a previously undescribed gene family.

An amino terminal signal sequence for membrane insertion (discussed in section 3.4) and the conserved heme binding motif FXXGXXXCXG common to all cytochrome P450 enzymes (Nelson *et al.*, 1993) can be clearly identified in the predicted protein (figures 3.19, 3.20 and 3.21). The presence of a tyrosine residue at the beginning of the heme binding motif (position 439) is extremely unusual as this residue is conserved as phenylalanine in all cytochrome P450s characterised to date, fewer than 1.5% of which diverge from the established motif at even a single residue (figure 3.21; Nelson *et al.*, 1993; Yu *et al.*, 1997). The substitution of tyrosine for phenylalanine results from a single transversion in the codon from UUU/C (Phe) to UAU/C (Tyr). At the protein level phenylalanine is an extremely hydrophobic amino acid, while the presence of a polar hydroxyl group on its side chain (the only structural difference between these two aromatic amino acids) makes tyrosine much less hydrophobic (Mathews & van Holde, 1990). The heme-binding site of *ORF2* also contains an arginine residue which is conserved in all filamentous fungal and most other P450s (figure 3.21).

3.3.7 An open reading frame highly similar to ORF1 can be identified 5' of *MSAS* in *Aspergillus terreus.*

The sequence for *A. terreus MSAS* submitted to the EMBL/GenBank/DDBJ databases (Pazoutova *et al.*, 1995) fortuitously includes 2 kb of sequence 5' to the coding region of the gene. A large open reading frame can be identified in this 5' region between nucleotides 643 and 29 of the database sequence which is conserved with *P. patulum ORF1* in both nucleotide and predicted amino acid sequences and is assumed to represent the presence of a homologue of this gene in *A. terreus*. The conserved open reading frames observed in these two species did not begin with ATG codons, suggesting the presence of an intron in the *ORF1* gene. Comparison of the regions upstream of *ORF1* in *A. terreus* and *P. patulum* revealed further similarity in a different reading frame and between the two regions of homology the position of an intron could be deduced (figure 3.23). The 5' splice site consensus element for this putative intron is GTAATC, the 3' splice site is CAG and an internal consensus of CGCTGAC is situated 18 bp from the 3' end. Protein products for

Cytochrome P450 gene	Heme-binding motif	Position
Rat CYP2J3	FSMGKRACLG	441-450
Phaseolus aureus CYP73	FGVGRRSCPG	440-449
Aspergillus parasiticus CYP60A1	FSIGPRNCIG	429-438
A. nidulans CYP60A2	FEFGPRSCIG	443-452
A. nidulans CYP59	FEFGPRSCIG	427-436
Fusarium. sporotrichioides CYP58	FSQGSRQCIG	449-458
Nectria. haematococca CYP57	FGAGSRSCIG	451-460
A. niger CYP53	FSTGPRACVG	454-463
Penicillium patulum ORF2	YGAGRRICPG	439-448

Figure 3.21 Alignment of amino acid sequences in the heme-binding active site of mammalian, plant and fungal cytochrome P450 enzymes.

Conserved amino acids are shown in bold type and the heme-binding cysteine is underlined. GenBank/EMBL accession numbers: CYP2J3, U39943; CYP73, L07634; CYP60A1, U62774; CYP60A2, U34740; CYP59, U34740; CYP58, U22462; CYP57, X73145; CYP53, X52521.

```
  1 .........MAPFVPYHYSAGQSTIVKFGGLLTTEFLEPPPGRCFLFRQT 41
              ||||:| ||||  | ||||||| ||| ||||| |: |||
  1 MGELFKWPSFSPFVPFHVSAGQTVIRKFGGLLTCEFLPPPPGRSFMMRQT 50
            .         .         .         .         .
 42 YRHTVEGSIPDNLRKLINSPDRPKGPPPHFHQFQTEYFRVESGVLGISVD 91
    ||| ||| ||:|||||| |  || ||| |||| |||||:|| |:  : |
 51 YRHSVEGPIPENLRKLIESDHRPDGPPMHFHQWQTEYFKVEEGICVVEVN 100
            .         .         .         .         .
 92 GVVRRITPEDGEISVKAGSVHNFFIHPDSPESMTVYLSASDSGNDYQLDR 141
    |    :||:| ||| ||| :| ||||||| | ||| ||||||| ||||||
101 GKQTMLTPDDEEISCKAGNIHRFFIHPDSRERMTVILSASDSGVDYQLDR 150
            .         .         .         .         .
142 VFFENWYGYWHDALLHDGGIDWIQFLAVSLSSQCIISLPIADMVVDPRRR 191
    ||||||||||||||: ||:| || | ||
151 VFFENWYGYWHDALLYQGGLDFIQTLCVSPQFLYF*.............. 186
```

Figure 3.22 Conservation of ORF1 predicted amino acid sequence in *Aspergillus terreus*.

Upper sequence is predicted amino acid sequence for *Penicillium patulum ORF1*, lower sequence is predicted amino acid sequence for *Aspergillus terreus ORF1*, identified in sequence 5' of *A. terreus MSAS* (Pazoutova *et al.*, 1995). Vertical lines indicate identity between amino acid residues, two dots indicate conserved amino acid similarities.

ORF1 in both species have been predicted and these two proteins are homologous along their lengths (figure 3.22).

Clustering of the gene represented by *ORF1* with *MSAS* in these two closely related fungi may simply be a consequence of a more general synteny, but the conservation of the proximity between *ORF1* and *MSAS* in two different organisms adds weight to the case for them being part of a secondary metabolic pathway gene cluster.

3.3.8 Transcriptional control motifs in *ORF1*.

The predicted promoter region for *ORF1* was examined for likely transcriptional control motifs. No pyrimidine rich region resembling a CT box could be identified in the region adjacent to the putative start of translation. A TATA-like motif (TATATA) is present 124 bp before the start of translation (figure 3.23, 384-389 bp). Eight GATA sites are present in the region between the 5' end of *ORF1* and the 5' end of *MSAS,* however none of these show the correct spacing of 3-40 bp. Five putative metal response elements can be identified between *ORF1* and *MSAS*; as in *ORF2* these all differ by a single nucleotide from the core consensus sequence (TGCRCNC).

3.3.9 Functional characterisation of *ORF2*.

A P450-like gene (*ORF2*) linked to *MSAS* has been isolated; it is now necessary to ascertain whether it has a role in the patulin biosynthetic pathway. Confirmation of this could be obtained by disruption of the chromosomal copy of the gene in *P. patulum* and analysis of any resulting changes in production of the patulin pathway intermediates. This approach was not considered to be practicable because although transformation systems based on the use of electroporation (Chakraborty *et al.*, 1991) and polyethylene glycol (Kiuchi *et al.*, 1991) have been developed for *P. patulum,* integration of the transforming plasmid has been predominantly ectopic. The four *E. coli* clones expressing the full length GST-ORF2 fusion protein (figure 3.18) are currently being assayed for enzymic function capable of catalysing selected reactions in the patulin biosynthetic pathway. Five reactions in the patulin biosynthetic pathway are potentially catalysed by cytochrome P450-type enzymes, as they involve the insertion of a single atom of oxygen into the substrate and are NADPH-dependent. Of these enzymes (indicated with an asterisk in figure 3.1), *m*-cresol methyl hydroxylase (3), *m*-hydroxybenzyl alcohol 6-hydroxylase (9) and *m*-cresol 6-hydroxylase (8) have previously been proposed as cytochrome P450s (Murphy *et al.*, 1974;

Murphy Lynen, 1975). Functional equivalence of the *ORF2* gene product with any of these enzymes will confirm the involvement of *ORF2* in the patulin biosynthetic pathway and the linkage of pathway genes. Functional eukaryotic cytochrome P450s, including the soluble fungal P450 CYP55, have been successfully expressed in *E. coli*, although high level expression required modification of the N-terminal codons (Barnes *et al.*, 1991; Okamoto *et al.*, 1997; Porter & Larson, 1991). *ORF2* was expressed as a GST fusion because it was felt that the fusion to GST might prevent the protein from being expressed in a membrane-bound form, ensure a high level of expression and facilitate purification of the ORF2 protein. If no appropriate activity can be detected in crude *E. coli* extracts, this could be a consequence of the fusion to GST affecting enzymic function in which case the fusion protein can be affinity purified using a glutathione sepharose matrix and GST removed by thrombin cleavage.

Figure 3.23 DNA sequence of cloned fragments located upstream of *MSAS* with the predicted protein sequences for ORF1 and ORF2 (cDNA).

Introns in *ORF2* and putative intron in *ORF1* are indicated by underlining. TATA and polyadenylation signal consensus sequences are indicated by bold underlining. GATA sites are indicated by double under- or over-lining depending on orientation. Metal Response Element consensus sequences in 5'-3' and 3'-5' orientations are highlighted in bold type. Possible CT boxes are shown in grey.

```
  1  AAGCTTCGCAAGATTAATCTCACTGACGGCCAACGGGAATAACCGACGAGTATTGACTTC   60

 61  CCGCCCGTAAGCAGTGACAAGTGGCGATCATGAAAGTGTGTTTAGGGATTTTGAGACGTT  120

121  GAGGACAAGCCGCCGTCAGCCCTAAGCCTGAATGGGATGCATGCTGATTAACCCAAAGGA  180

181  GCGCCGGGGATGATTTCTTCAAAGGGTGCCCGTTTAACTGTCTCACCGTTTAGGAACCTT  240

241  ACAGGAACTCCGTAAACATACTCCGGAGCGAATGTGTATCTAAATATACTCATTTTGATC  300

301  AGGCCATTTCGACATGCCTTGAAAGGCAGTCCATAGAGGACTGGACTCTACATGTGAGCA  360

361  TAGCAAGGAATCAGGATATCAAGTATATAGGTAGTAGCAGGCGCATCTGGAGTATTTCTA  420

421  GATATCTTTAGGAGAAGCGCTCAAATATTCATGCTTTTGTTGTAAAAATCCGACTTCGAA  480

481  CAGCTAGCCAACCATCTGAAGTGTCTTGTTACAATGGCCCCATTTGTGCCCTACCACTAC  540
                                    MetAlaProPheValProTyrHisTyr
                                    ORF1 ──►

541  TCAGCTGGCCAGTCGACTATTGTCAAGTAATCAATCCCCCTTGTCCTTCTGTATATTAAA  600
     SerAlaGlyGlnSerThrIleValLy

601  TAAAAACTCTATTGGTCGCTGACGTAGACCCAATTTTGCAGATTTGGTGGCCTTCTGACC  660
                                          sPheGlyGlyLeuLeuThr

661  ACCGAATTTCTCGAACCTCCACCAGGCCGTTGCTTTCTGTTCCGACAGACATACCGCCAC  720
     ThrGluPheLeuGluProProProGlyArgCysPheLeuPheArgGlnThrTyrArgHis

721  ACAGTTGAAGGATCAATTCCGGATAATCTGCGCAAGCTCATCAACAGCCCTGATCGACCA  780
     ThrValGluGlySerIleProAspAsnLeuArgLysLeuIleAsnSerProAspArgPro

781  AAGGGTCCGCCGCCGCATTTCCATCAGTTTCAAACCGAGTACTTCCGTGTCGAGAGTGGA  840
     LysGlyProProProHisPheHisGlnPheGlnThrGluTyrPheArgValGluSerGly

841  GTGCTCGGGATTTCAGTGGATGGAGTTGTCCGCCGCATCACCCCGGAAGATGGAGAAATC  900
     ValLeuGlyIleSerValAspGlyValValArgArgIleThrProGluAspGlyGluIle

901  TCGGTGAAGGCAGGCAGTGTGCATAACTTCTTCATTCATCCAGATTCCCCCGAGAGCATG  960
     SerValLysAlaGlySerValHisAsnPhePheIleHisProAspSerProGluSerMet

961  ACGGTCTACTTGAGTGCATCTGATTCTGGAAACGACTACCAGCTCGATCGGGTTTTCTTT 1020
     ThrValTyrLeuSerAlaSerAspSerGlyAsnAspTyrGlnLeuAspArgValPhePhe

1021 GAAAACTGGTACGGCTATTGGCACGATGCATTGTTGCATGATGGAGGTATCGATTGGATC 1080
     GluAsnTrpTyrGlyTyrTrpHisAspAlaLeuLeuHisAspGlyGlyIleAspTrpIle

1081 CAGTTCCTTGCAGTAAGTCTCTCTTCTCAGTGCATAATTTCCTTGCCAATAGCTGACATG 1140
     GlnPheLeuAlaValSerLeuSerSerGlnCysIleIleSerLeuProIleAlaAspMet

1141 GTAGTAGATCCAAGACGGCGGGGATGCCTATACACCCGCTCCCGCATGGTGCCCTTCCGT 1200
     ValValAspProArgArgArgGlyCysLeuTyrThrArgSerArgMetValProPheArg

1201 CGGCAGGTCGGATACTGGACCTGTGTAATCGTGGGCCGATGGATTGGCGGACTCCTAGGG 1260
     ArgGlnValGlyTyrTrpThrCysValIleValGlyArgTrpIleGlyGlyLeuLeuGly

1261 TACAAGCCCTTCTTCCGAGAATATACTACGGATTGGGACTTTGCCGTTGCCAAGATGAAG 1320
     TyrLysProPhePheArgGluTyrThrThrAspTrpAspPheAlaValAlaLysMetLys

1321 GGTTCCTTCTTCCAGCGACATCTGGTGCATGCTGCATTCGAGGAAGAGAAACCCTGGGCG 1380
     GlySerPhePheGlnArgHisLeuValHisAlaAlaPheGluGluGluLysProTrpAla

1381 AAACAGGCAGAGCTGGAACCCAAGGCTAAGCCGGAAAATGCCGAGTTTGAGCCGTGGGTT 1440
     LysGlnAlaGluLeuGluProLysAlaLysProGluAsnAlaGluPheGluProTrpVal

1441 GAGGATATGTCCCCCACACCATTGGCATTGGGTCCAGTGGCCTACGAGGAGGGGCAGTTC 1500
     GluAspMetSerProThrProLeuAlaLeuGlyProValAlaTyrGluGluGlyGlnPhe

1501 AAAGGCGCACAGGCCAAGTCGCTGAATGGGTCCAACGGTCACTCCACAGGGGTGGAAAGG 1560
     LysGlyAlaGlnAlaLysSerLeuAsnGlySerAsnGlyHisSerThrGlyValGluArg
```

```
1561  AAGTTGGAGCAGTTGGGGTCTTCGACGCAGCGACGAACAGTGGTCAACAATGCGGACAAG   1620
      LysLeuGluGlnLeuGlySerSerThrGlnArgArgThrValValAsnAsnAlaAspLys

1621  TAGAGAGAGATATGTGACATATTTTCCAAGACTTGAGTTCGAACCTTATTTCCCTTGGCT   1680
      End

1681  TCCTCGTAGGGCAAATACAAAATGAACTTCCAATTGACCAAAACCCTGTTCAGGAAGAAG   1740

1741  GTTTGGGCAATGAAAAGAGTAAATCAGTGGTGACCAAAGGGGAGATAGCATGCCCCCGGG   1800

1801  TGTATTAATTGTCTTCTAAGTTGTATTGTATGATGACGCCATATTTGTGAAGGTCGAGCT   1860

1861  TATAGGTCTCATGGCAAAGCTCGTAGGGTCGTAGGGAAAATACAACGCCGAGGATTATTA   1920

1921  CAGATAGTAACAAGAAGGATGGGAATAATATGAATAGTAATAATAAGGAGAGCAAAGAGA   1980

1981  TACTAATAGTACTCAGGCTAGTTTCTGTTGCAGGCGAATCGCACAAAGTGGACGAGGATT   2040

2041  TGTGTAGTATTCATCTTCAACGACTAATTCTTGACCGGAGGGATCCATGGGGAGGCATTC   2100

2101  GCTTCCTTTTCTGCATGTCATGTCGAACTGATAGTTCCGTATAACTAGTGTAGTTGCCAC   2160

2161  CTGGACAATAATGACTCTCTGTAAGACAATGAAACTTTATATGCAATCCTGATGCGTCAA   2220

2221  AGTCAACAACCTCAGGATACTGTATCCTTTACATATAATTTTACGACACTACATATGGAT   2280

2281  GTAGTATTATGTATATTATACTAGAACATCCTAAGGACTTTCGAAAATGGTAATGCAAGT   2340

2341  TTACCCTTATTCTACGGCGTTGTCATTTATATTTATAAAGTCAAAGGTTATATATAGGAA   2400

2401  AGTCGTCTGCACAAATTAAGCGCCTTTCTTTTGGCCGTCAACCTACAAAACTGGCACAGC   2460

2461  TAAAGAGAATGTATCTGAGGAGACCACTGAAAAGGCTACTCAAAGGATATCACAAGGCCG   2520

2521  CTTATGCGGCCGACTATTCTTGTCCTCTTACATAGACTACCAGGTATAGACGGATCGACA   2580

2581  TCGTGCACTTGATGTCAATAGGGAATGTGAAACTGTCTATTCAGTTGAGGCTCCGATCGT   2640

2641  ACCTTACATCATTAAGGAGACATCATTGATATTCTCTCGGTTGACAGAATGGGGTGTGCC   2700

2701  CATAAATAATATATGCATACGTGAGATCTCTATACGAGCAGACTCCGATTCTCCTTTAGG   2760

2761  AATCTTAAACCATCGCCCACTTAGCTGATCCCCCATCTCGAAAGAGGTAGTATCCCCGCT   2820

2821  CCCCGCGTTGATGAGAACAATAAGTATGTATAGACAATATTTCTGCCCCGAGGAACAATT   2880

2881  TTTGGCCATTTGTCAACACTATGGATATCCTACAGTTCGCACCAACGCATCTTCTGGCAA   2940
                    MetAspIleLeuGlnPheAlaProThrHisLeuLeuAlaI
                    ORF2 ──►

2941  TACTGCTTAGCAGTACCAGCGTATTGTTCCTGGTCACGTATCTCCTTCGTGCAGGACATC   3000
      leLeuLeuSerSerThrSerValLeuPheLeuValThrTyrLeuLeuArgAlaGlyHisA

3001  GACCCTCAGAACTACCGGACGGTATGTGGTTGCAGTGACGAGCCAAGACTTCATATCTAA   3060
      rgProSerGluLeuProAspG                Intron I

3061  CTAGTTCACAGGTCCTCCAACCGTTCCACTTTTTGGAAATGAGCTGCAGGTGCCAAAATC   3120
                  lyProProThrValProLeuPheGlyAsnGluLeuGlnValProLysSe

3121  AGATGCACATTTCCAGTAAAGACCAACCACCCTCCGTGATGTATCAAGGCAATGTGCTAA   3180
      rAspAlaHisPheGl                      Intron II

3181  GTATTTCGTGTAGGTTCTCCAAGTGGGCCAAGCAATATGGCGGCTTTTTCACGCTCAAAC   3240
                    nPheSerLysTrpAlaLysGlnTyrGlyGlyPhePheThrLeuLysA

3241  GATACAACAACACCACCATTGTCATCAGTGACCAGAAACTCATCAAGCAACTGCTGGACA   3300
      rgTyrAsnAsnThrThrIleValIleSerAspGlnLysLeuIleLysGlnLeuLeuAspL

3301  AAAAAAGCAACATATATTCTCACCGACCAGCGTCCCTAGTATCGCACCTTATTACGCAGA   3360
      ysLysSerAsnIleTyrSerHisArgProAlaSerLeuValSerHisLeuIleThrGlnS

3361  GTGACCATTTGCTCGTTATGCAATATGGTGAGCGATGGCGCATGCTGCGTAAAACCATTC   3420
```

```
      erAspHisLeuLeuValMetGlnTyrGlyGluArgTrpArgMetLeuArgLysThrIleH

3421  ATCAATACTTCATGGAGCCCCGCTGCGAGCGTGAACACTGGAAAGTCCAAGAGGCTGAAG  3480
      isGlnTyrPheMetGluProArgCysGluArgGluHisTrpLysValGlnGluAlaGluA

3481  CCAAACAGATGTTGCACGACTTTCTGACCATGCCAGAGGATCACATGCTGCACCCTAAAC  3540
      laLysGlnMetLeuHisAspPheLeuThrMetProGluAspHisMetLeuHisProLysA

3541  GGTACAGTAACAGTATCACCAACTCTCTTGTCTTTGGTATTCGTACCAAAACTGTCCACG  3600
      rgTyrSerAsnSerIleThrAsnSerLeuValPheGlyIleArgThrLysThrValHisA

3601  ACGAGTATATGAACAAGCTTTTCTACCTGATGGACAAATGGTCATTGGTGCAGGAGCTAG  3660
      spGluTyrMetAsnLysLeuPheTyrLeuMetAspLysTrpSerLeuValGlnGluLeuG

3661  GCGCCACCCCACCTGTGGACTCGTTCGCTTTGCTGCGTTATGTCCCGCAGTGGCTGTTGG  3720
      lyAlaThrProProValAspSerPheAlaLeuLeuArgTyrValProGlnTrpLeuLeuG

3721  GTAACTGGCGCAATCGAGCGATCGAAGTGGGTGATCTAATGCAGTCGCTTTACAAGACCG  3780
      lyAsnTrpArgAsnArgAlaIleGluValGlyAspLeuMetGlnSerLeuTyrLysThrV

3781  TTCTAGATCAGGTTCGGGAGCGCCGCCAGCGGGGTATTCAGCGGGACTCCTTCATGGACC  3840
      alLeuAspGlnValArgGluArgArgGlnArgGlyIleGlnArgAspSerPheMetAspA

3841  GAGTACTAGATTCAATGAAACAGACCCCACTCACGGAGAATGAGTTGCGATTCCTAGGAG  3900
      rgValLeuAspSerMetLysGlnThrProLeuThrGluAsnGluLeuArgPheLeuGlyG

3901  GAGTTCTGATGGAAGGAGGGTCCGACACTTCCTCCTCGCTTATATTGACCATTATCCAAG  3960
      lyValLeuMetGluGlyGlySerAspThrSerSerSerLeuIleLeuThrIleIleGlnA

3961  CTATGACTAAATATCCAGAGGTGCAAGCGAAGTATGCCACTCGGATATTTTCTCCCCATC  4020
      laMetThrLysTyrProGluValGlnAlaLy

4021  TGTTTCAGGAAAAGCGCTAATGGAGCAATCTCACAGGGCCCATGCACAAATCGACTCCAT  4080
         Intron III                          sAlaHisAlaGlnIleAspSerIl

4081  TGTCGGCCAGGACCGCTCTCCAGCTTGGTCTGATTGGCCTCGGCTGCCCTATATAAACAT  4140
      eValGlyGlnAspArgSerProAlaTrpSerAspTrpProArgLeuProTyrIleAsnMe

4141  GATCATCAAGGAGTCCCACCGGTGGCGTCCCGTCAGCCCTTTGGGTGTTACTCACGCTGT  4200
      tIleIleLysGluSerHisArgTrpArgProValSerProLeuGlyValThrHisAlaVa

4201  CGCTGAAGGTCTGTCTTACAAAATCTTACAAATTTGAATGCATCTTTCGGCATTCCAGAC  4260
      lAlaGluA

4261  GCCGAAGACCGAGAGAAGAGCGAAAGAACACATTGGCTAATAATATTTTCTAGATGACCA  4320
                  Intron IV                                   spAspHi

4321  CATCGACGGAAAACTCATTCCGCAAGGATCGAGCATCGTTCTCAATGTCTGGGGCATGCA  4380
      sIleAspGlyLysLeuIleProGlnGlySerSerIleValLeuAsnValTrpGlyMetHi

4381  TCATGATAGTGATCGATGGCAGGAACCAGAGCATTTCCAGCCTGAACGTTTTGCTGATTT  4440
      sHisAspSerAspArgTrpGlnGluProGluHisPheGlnProGluArgPheAlaAspPh

4441  CCCGGCTTTAGCCTCGACCTACGCCGGCTCCGAGCGGAGGGATCATATTGGTTACGGAGC  4500
      eProAlaLeuAlaSerThrTyrAlaGlySerGluArgArgAspHisIleGlyTyrGlyAl

4501  GGGCCGTCGAATTTGCCCAGGCATCCACTTAGCGGAGCGAAACTTGATCATAGGCATTGC  4560
      aGlyArgArgIleCysProGlyIleHisLeuAlaGluArgAsnLeuIleIleGlyIleAl

4561  CAAGTTGCTCTGGGCGTTCGAGTTTGTGGAGCCACCTGGCAGTGATAGTGACATTTCGGC  4620
      aLysLeuLeuTrpAlaPheGluPheValGluProProGlySerAspSerAspIleSerAl

4621  CCACTCTGGAGCTAGCAAAGGCTTCCTCCATTGCCCCAAGGACTACGGATGTGCCATTCG  4680
      aHisSerGlyAlaSerLysGlyPheLeuHisCysProLysAspTyrGlyCysAlaIleAr

4681  CCTCCGTTCTCCGGAAAAGAGAGCGACCATTATGCGAGAATTTGCCGAGGCACAGGAGGT  4740
      gLeuArgSerProGluLysArgAlaThrIleMetArgGluPheAlaGluAlaGlnGluVa

4741  GTTCGCTCAATTTGATTAGCTAGATTAGAGAATAAGCACTAGAATACGATTGACCCCTCC  4800
      lPheAlaGlnPheAspEnd

4801  ATTGGGGGAAAGTTGTTTGAGGGTCCATCGAAAGCACGTATATCCACTCCATGTACGTAC  4860
```

```
4861  ATGTCAACGGCACTAAAACCCTATTTATTTGCAGATCATCCGTAATAGTGGATACATTTA  4920
4921  ATGTAAATGAAGTCCACGACTGGTGTAGCTTGCTTCGTAATTGTTCATTTATTGCCTCAA  4980
4981  TTAAACTCAGTACTTTGCTCATCAAGGAGTGTAAGATTTTAGTAGCAAGAAAGGCAGCAA  5040
5041  TGAAGGTCCTCATCCTCCATACTAGCAACTGAATCCCGACAAATTGGGCCGTTTGCCATG  5100
5101  GAATAAAAATTTGATCCCTGCAATCAATCAATGCACAACGTATTTTGGAATCACAAACGG  5160
5161  GCAACTCCGCATCTGCAATATATGCATGTCGCCTAGAGGAGTAGCGGAACAGCGGTCCTC  5220
5221  ACTCCGATTTCTTCACAGATTAAAGCTAATTTCTGAGATAAGGTGATCTACTCAGCACGC  5280
5281  CTTCATGCTTGATTGGAGTTCACGGAGTCTCTGAAAAGGCGTGGGGGTGGATAACTTTTT  5340
5341  GTGAAGAATATCCCTGCACGGCCGGAAGATATTATGGGATCCGAGCTC               5388
```

CHAPTER 4

PCR-DERIVED PROBES FOR FUNGAL POLYKETIDE SYNTHASE GENES

4.1 INTRODUCTION

4.1.1 Cloning genes from filamentous fungi

The various strategies devised to clone chromosomal genes from filamentous fungi have been reviewed by Turner (1991) and are briefly described below.

Early strategies for cloning fungal genes involved complementation of allelic mutations in *E. coli* (e.g. Kinghorn & Hawkins, 1982) or yeast (e.g. Berse *et al.*, 1983). The development of fungal transformation systems (reviewed by van den Hondel & Punt, 1991) allowed homologous expression of the gene of interest and identification by complementation of a mutant. This approach avoided the problems of intron splicing and promoter recognition encountered using bacterial and yeast systems. The low transformation efficiencies achieved in many fungal species (e.g. 3-5 stable transformants/μg DNA in the case of *Verticillium dahliae* [Dobinson, 1995]) have necessitated the use of heterologous systems, which are of limited use. More recent developments include a method for rapid complementation analysis of mutant filamentous fungi by co-transformation of digested genomic DNA from a donor species with an autonomously replicating vector (Bowyer *et al.*, 1994; Gems *et al.*, 1994; Verdoes *et al.*, 1994). This technique has generally used a genetically well characterised species such as *Aspergillus nidulans*, *A. niger* or *Neurospora crassa* as the mutant recipient. Genes have been isolated from fungal genetic libraries by the use of techniques including hybridisation of degenerate (mixed sequence) oligonucleotides designed to bind to regions coding for conserved amino acid sequence, hybridisation of DNA probes derived from highly homologous genes and immunological screening of cDNA expression libraries. Where the gene product possesses regions of conservation with known sequences, PCR amplification employing degenerate primers can provide gene-specific DNA probes for library screening or circumvent the lengthy process of constructing and screening gene libraries by amplifying the requisite regions of the gene (Compton, 1990; Gould *et al.*, 1989; Lee & Caskey, 1990). This procedure, which originally used a cDNA template, is sometimes designated MOPAC (mixed oligonucleotide primed amplification of cDNA (Lee & Caskey, 1990; Lee *et al.*, 1988). The use of degenerate primers was subsequently successfully extended to genomic templates (Gould *et al.*, 1989). Genes encoding enzymes have also been cloned from a filamentous fungus by expression of a cDNA library in

Saccharomyces cerevisiae followed by assaying the library clones for the appropriate enzyme activity (Dalboge and Heldthansen, 1994).

Subtractive and differential hybridisations are methods for screening a genomic or cDNA library for genes which are expressed under one particular set of conditions, but not another (eg. Ehrenshaft and Upchurch, 1991). This can narrow down the number of clones that need to be screened to find the gene of interest.

4.1.2 Cloning fungal polyketide synthase genes.

The methods used to clone fungal PKS genes reported in the published literature are summarised in section 1.4.2. Heterologous probing, probably the simplest method available for cloning a gene, was used only in the cloning of *MSAS* from *A. terreus* (Fujii *et al.*, 1996) suggesting that PKS genes which are not functional homologues are too highly divergent in sequence. This contrasts with the type II streptomyces PKS genes (Malpartida *et al.,* 1987) but is similar to the situation with the streptomyces type I PKS genes (Blanco *et al.*, 1993). In addition, two fungal PKS genes have been cloned in this laboratory. A genomic fragment (2Y) containing the ketosynthase domain of a putative novel PKS gene from *Aspergillus parasiticus* was isolated from a genomic library by heterologous hybridisation, using the condensing domain probe CON1 (figure 3.2) from the *Penicillium patulum* gene *MSAS* (Walsh, 1993). Several screening attempts were required before a single weakly-hybridising plaque was identified by the heterologous probe. In contrast, subsequent use of a subcloned restriction fragment of clone 2Y as a homologous probe rapidly identified numerous overlapping clones in an amplified genomic library. The complete coding sequence of the putative PKS gene, designated *APPKS1,* shows a high degree of sequence conservation with *MSAS* and may in fact be a homologue of this gene. While bands hybridising strongly to the *MSAS* ketosynthase and reductase domain probes and to a *Streptomyces* condensing and ACP domain probe (*actI*) were clearly identifiable on Southern blots of *A. parasiticus* genomic DNA, these did not correspond to the cloned gene (Walsh, 1993). These results suggested that heterologous probing was not a generally applicable method for identifying novel PKS genes and an alternative approach using PCR was investigated. Degenerate primers were designed using regions of continuous amino-acid conservation between *APPKS1* and *MSAS* (the only fungal PKS sequences available at the time) and a cloned PCR product obtained using these primers was used to probe a *P.*

patulum genomic library. A putative second *P. patulum* PKS gene, homologous to the cloned PCR product, was isolated by this method, as described in Chapter 5.

4.1.3 Uses for novel fungal polyketide synthase genes

There are several reasons why the ability to clone novel fungal PKS genes would be valuable. As discussed in section 1.4.4, fungal PKS enzymes are thought to possess several unusual features not found in the bacterial PKS and these features could be exploited in the generation of novel polyketides via combinatorial biosynthesis. Fungal PKS enzymes are particularly suitable for heterologous gene expression (Bedford *et al.*, 1995) and therefore the cloning of novel PKS genes may facilitate the industrial production of useful polyketides. Lastly there are instances where it may be desirable to disrupt biosynthesis of a harmful polyketide metabolite, for example to generate nontoxigenic biocontrol strains of mycotoxin-producing fungi. Cloning and characterisation of the PKS genes involved could represent a step towards achieving this goal (Linz & Pestka, 1992).

4.2 RESULTS

4.2.1 PCR primer design

Primer design was based on the four available fungal PKS sequences with the aim of extending the range of fungal PKS genes for which homologous probes could be obtained. The PKS sequences were observed to fall into two subgroups on the basis of amino acid sequence conservation in functional domains; one subgroup consisted of *A. nidulans* wA and *C. lagenarium* PKS1, the other included *P. patulum* 6MSAS and *A. parasiticus* APPKS1.

Two pairs of degenerate PCR primers of 20-21 nucleotides in length were designed to bind to regions on either side of the condensing domain, where amino acid sequence was conserved within these subgroups but not between them (figure 4.1). Sequence identity for the regions bounded by the LC primers in these four gene sequences has a mean of 63.5% within groups but only 46.5% between groups (GAP program, Devereux *et al.*, 1984). Degeneracies for these primers range from 128 to 1024 fold; these are relatively low values as primers with degeneracies of greater than 10^5 have been used successfully in PCR (Gould *et al.*, 1989), and should provide highly specific primers. The use of the universal base inosine at positions of four-fold base redundancy kept the complexity of the primer mixtures to a minimum.

To facilitate cloning and sequencing of amplified products from the LC1/2c primers, longer versions of these primers were synthesised with non-degenerate 5' 'tails'. These tails are bifunctional: they allow the annealing of sequencing primers complementary to the tail sequence on the amplified product and thereby facilitate direct sequencing of the PCR product; also they contain restriction sites to make 'sticky-ended' cloning of the product possible. The sequences for the extensions to LC1 and LC2c were chosen from oligonucleotides already available in the laboratory to contain suitable restriction sites and to be suitable for use as sequencing primers (figure 4.2). The oligonucleotide tailing LC1 was designated BAMECO1 (GATCGTTGGATCCTCTAG) and that tailing LC2c was designated STREP1T (TAAGATCTCGAGCTCTAGAG).

B.

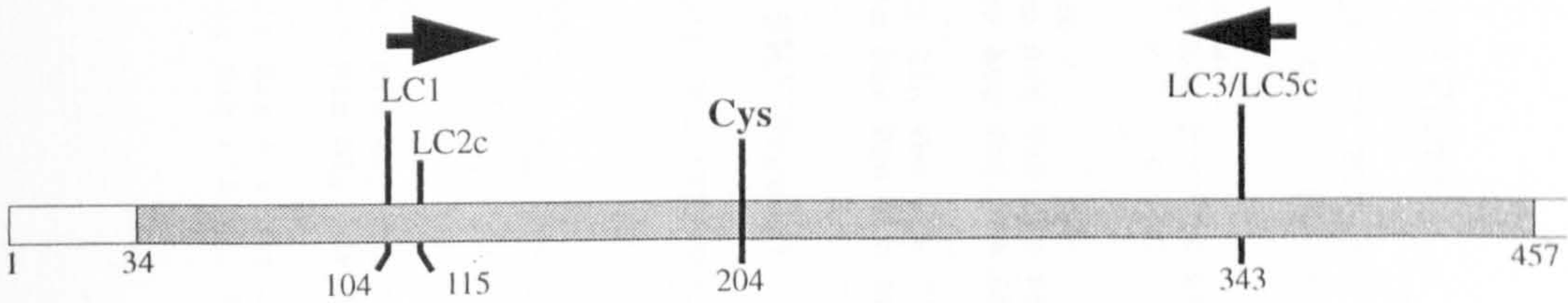


Figure 4.1 Design of LC-series degenerate PCR primers.

A (following page). DNA sequences are shown for each primer, with the degeneracy of the primer indicated to the right. Beneath is the genomic DNA sequence that the primer is designed to bind to, together with the amino acid sequences from which the primers were designed. The universal base inosine is indicated in the primer sequence by the letter I.

B (above). This diagram shows the amino-terminal end of the MSAS coding region (empty box) with the extent of the ketosynthase domain (as defined by Donadio & Katz, 1992) shown by the stippled box. The positions of the region encoded by the LC-series primer binding sites and the active site cysteine are indicated, with amino acid numbers shown below the diagram.

A.

Lewcon1 (5' ßKS Primer)

```
GAT CCI AGI TTT TTT AAT ATG      Degeneracy=32
  C     C     C   C   C
```

Genomic DNA sequence

```
GAT CCN AGN TTT TTT AAT ATG
  C     C     C   C   C
```

Amino Acid sequence

Asp Pro Arg Phe Phe Asn Met	*A. nidulans* wA
Asp Pro Arg Phe Phe Asn Met	*C. largenarium* PKS1
Asp Cys Gln Phe Phe Gly Ile	*P. patulum* MSAS
Asp Ala Ser Phe Phe Asn Ile	*A. parasiticus* APPKS1

Lewcon2c (3' ßKS Primer)

```
GTI CCI GTI CCG TGC ATT TC Degeneracy=4
                  A       C
```

Genomic DNA sequence

```
GAA ATG CAC GGN ACN GGN AC
  G       T
```

Amino Acid sequence

Glu Met His Gly Thr Gly Thr	*A. nidulans* wA
Glu Met His Gly Thr Gly Thr	*C. largenarium* PKS1
Glu Ala His Ala Thr Ser Thr	*P. patulum* MSAS
Glu Ala His Ala Thr Ser Thr	*A. parasiticus* APPKS1

Lewcon3 (5' ßKS Primer)

```
GCI GAA CAA ATG GAT CCI CA Degeneracy=8
      G   G       C
```

Genomic DNA sequence

```
GCN GAA CAA ATG GAT CCN CA
      G   G       C
```

Amino Acid sequence

Ala Glu Gln Met Asp Pro Gln	*P. patulum* MSAS
Ala Glu Gln Met Asp Pro Gln	*A. parasiticus* APPKS1
Ala Leu Gln Ala Asp Pro Ala	*A. nidulans* wA
Ala Phe Gln Thr Asp Pro Met	*C. largenarium* PKS1

Lewcon 5c (3' ßKS Primer)

```
GTI GAI GTI GCG TGI GCT TC Degeneracy=4
                  A       C
```

Genomic DNA Sequence

```
GAA GCN CAC GCN ACN TCN AC
  G       T
```

Amino Acid sequence

Glu Ala His Ala Thr Ser Thr	*P. patulum* MSAS
Glu Ala His Ala Thr Ser Thr	*A. parasiticus* APPKS1
Glu Met His Gly Thr Gly Thr	*A. nidulans* wA
Glu Met His Gly Thr Gly Thr	*C. largenarium* PKS1

LC1EXT

*Bam*HI

GATCGTTGGATCCTCTAGATCCIAGITTTTTTAATATG

C C C C C

LC2cEXT

*Bgl*II *Sac*I

TAAGATCTCGAGCTCTAGAGTICCIGTICCGTGCATTTC

A C

*Xho*I *Xba*I

Figure 4.2 Tailed degenerate PCR primers for sequencing and cloning of amplified PKS fragments. Regions corresponding to sequencing primers BAMECO1 (LC1EXT) and STREP1T (LC2cEXT) are highlighted in bold type. Restriction sites are indicated by lines. The universal base inosine is indicated in the primer sequence by the letter I.

4.2.2 Two sets of PCR products are obtained from deuteromycete genomic DNA using the LC1/2c and LC3/5c primer pairs.

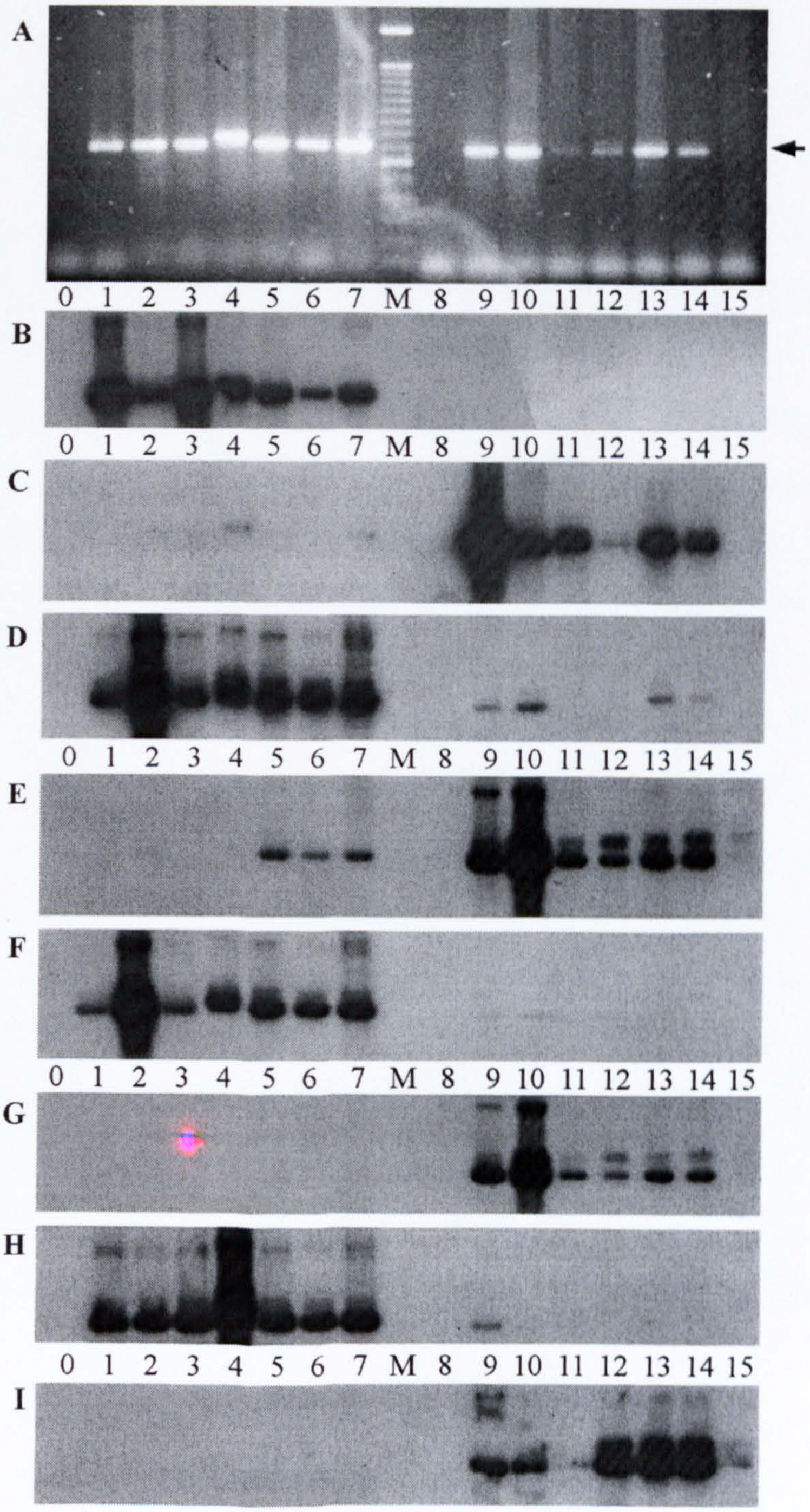
PCR products were obtained from a range of deuteromycete fungal genomes with both sets of LC primers (figure 4.3 A). Template DNA concentration was not standardised; a fixed amount of undiluted genomic DNA preparation was added to each reaction, resulting in template concentrations of 6.4-33.2 ng/ul. PCR products of approximately 700-800 bp in length were obtained with both pair of primers from every genome except that of *Drechslera monoceras,* from which the LC3/5c primers failed to amplify a product. Although present in most cases as a single band on electrophoresis (figure 4.3 A), each product may be composed of a mixture of comigrating fragments amplified from more than one target sequence.

To ascertain whether these products were indeed derived from structurally related PKS genes, they were blotted onto nylon membranes and probed with ^{32}P-labelled *P. patulum*, *Cytospora* sp. and *A. parasiticus* PCR products from each primer-pair (figure 4.3, B-I). The results show that, in general, products from one primer-pair hybridise strongly to each other, but not to products from the other primer-pair. Where cross hybridisation between PCR products from different primer pairs was observed (figure 4.3, E) this could be

Figure 4.3 Two sets of PCR products were obtained from deuteromycete genomic DNA using LC1/2c and LC3/5c primer pairs.

A. PCR products of approximately 700-800 bp (arrowed) were obtained using genomic DNA templates from a range of deuteromycete fungi with the two primer-pairs, LC1/2c (lanes **0-7**) and LC3/5c (lanes **8-15**). Thermal cycling parameters for all reactions were: (94° C, 3') x 1, (94° C, 1'; 55° C, 1'; 72° C, 3') x 34, (72° C, 10') x 1. PCR primers were used at the following concentrations: LC1, 3.2 µM; LC2c, 0.4 µM; LC3 1.2 µM; LC5c, 0.8 µM. Templates were (in 25 µl reaction volume): **0** & **8**, no DNA (negative control); **1** & **9**, *Penicillium patulum* (230 ng); **2** & **10**, *Cytospora* sp. (830 ng); **3** & **11**, *Penicillium citrinum* (160 ng); **4** & **12**, *Aspergillus parasiticus* (550 ng); **5** & **13**, *Phoma* sp. C2932 (380 ng); **6** & **14**, *Phoma etheridgii* (260 ng); **7** & **15**, *Drechslera monoceras* (560 ng). Reaction products were fractionated on a 1.1% agarose gel in TAE buffer at 110 mA. **M**, 100 bp ladder DNA size markers (Gibco-BRL).

B.-I. Hybridisations of PCR products from LC1/2c and LC3/5c primer pairs. PCR products shown in **A** were blotted from duplicate gels onto nylon membranes and probed with ^{32}P-labelled PCR products from: **B-C**, *P. patulum* (= part **A**, lanes **1** and **9**); **D-G**, *Cytospora* sp. (= part **A**, lanes **2** and **10**); **H-I**, *A. parasiticus* (= part **A**, lanes **4** and **12**). **B**, **D**, **F**, **H** were probed with LC1/2c PCR product; **C**, **E**, **G**, **I** were probed with LC3/5c PCR product. All hybridisations were carried out at 50° C, with washing at: **B** & **C**, 50° C; **D** & **E**, 55° C; **F** & **G**, 65° C; **H** & **I**, 55° C.


A
0 1 2 3 4 5 6 7 M 8 9 10 11 12 13 14 15
B
0 1 2 3 4 5 6 7 M 8 9 10 11 12 13 14 15
C
D
0 1 2 3 4 5 6 7 M 8 9 10 11 12 13 14 15
E
F
0 1 2 3 4 5 6 7 M 8 9 10 11 12 13 14 15
G
H
0 1 2 3 4 5 6 7 M 8 9 10 11 12 13 14 15
I

removed by increasing the temperature of the post-hybridisation washing step and thereby increasing its stringency (figure 4.3, G).

In several cases strong hybridisations to PCR products not distinguishable on the stained agarose gel, of approximately 800-900 bp and 1.4-1.5 kb are apparent. These products were assumed to be artefacts of the PCR and were not investigated further. The size of the 1.4-1.5 kb products suggests that these may be dimers of the 700-800 bp products. The products migrating consistently with a size of 800-900 bp are probably single stranded copies of the target region arising from asymmetric PCR occurring due to unequal primer concentrations (Gyllensten and Erlich, 1988). Single stranded DNA would migrate through the agarose gel at a different rate to double stranded DNA and would also fluoresce much less after ethidium bromide staining (McCabe, 1989).

Products of two different sizes were obtained in PCR using the LC1/2c primers with *A. parasiticus* template DNA (figure 4.3A, lane 4). The minor product is approximately the same size as LC1/2c products from other species while the major product appears to be about 50 bp larger. Both bands were labelled and used to probe the other deuteromycete PCR products (figure 4.3, H). Subsequently the two *A. parasiticus* LC1/2c products were isolated separately by excision from an agarose gel and re-amplification of the minor product, labelled and used to probe the set of LC1/2c and LC3/5c PCR products (figure 4.4). Differences in hybridisation patterns of the major and minor *A. parasiticus* products are clearly apparent, although both only hybridise to other products of the LC1/2c primer pair. The hybridisation pattern of the major product (figure 4.4, C) is almost identical to that seen for the mixed products (figure 4.3, H).

A PCR product of approximately 750 bp was amplified from the genome of another deuteromycete, *Verticillium dahliae,* using the LC1/2c primer pair (figure 4.5). No product was obtained from *V. dahliae* using the LC3/5c primer pair. Dilution of the genomic DNA was required to obtain a product, indicating the presence of PCR-inhibiting substances in the original preparation (figure 4.5).

4.2.3 Hybridisations of PCR products to genomic Southern blots.

Genomic DNA from *P. patulum* and *A. parasiticus* was digested with *Bam*HI and *Eco*RI restriction enzymes (figure 4.6) and blotted onto a nylon membrane. LC1/2c and LC3/5c PCR products from both genomes were ^{32}P-labelled and used to probe the Southern blots

(figures 4.7 and 4.8). Each PCR product hybridised to one or more restriction fragments in each genomic digest and showed different patterns of hybridisation to each fungal genome (figures 4.7 and 4.8).

The 5.2 kb *Eco*RI and 1.3 kb *Bam*HI fragments seen in the *P. patulum* genomic digest probed with the *P. patulum* LC3/5c PCR product (figure 4.7, B) are the expected hybridisations to known restriction fragments of *MSAS* (Beck *et al.,* 1990, Walsh, 1993). The strong hybridisations to 15.1 kb *Eco*RI and 2.0 kb *Bam*HI fragments indicate that the LC3/5c primer pair has amplified at least one additional product from the *P. patulum* genome. These fragment sizes are consistent with hybridisation of this probe to the *MSAS*-like PKS gene *PKS2* which was cloned from the *P. patulum* genome (see chapter 5). The 5.2 and 4.2 kb *Bam*HI and the 12.9 and 4.5 kb *Eco*RI hybridising bands in the *A. parasiticus* genomic digest probed with the homologous LC3/5c PCR product (figure 4.8, C) may be accounted for by the expected hybridisations to *APPKS1* (Walsh, 1993).

The LC1/2c PCR products obtained from *P. patulum* and *A. parasiticus* hybridised strongly to restriction fragments which did not correspond to any previously described PKS genes. Cloning and DNA sequencing the *A. parasiticus* LC1/2c PCR products confirmed that they were derived from novel PKS genes.

4.2.4 DNA sequencing of LC1/2c PCR products.

During the PCR process, errors may be introduced into the nucleotide sequence of the product. For *Taq* DNA polymerase these are mainly T→C transitions and other single base substitutions, although single base deletion errors may also occur (Eckert & Kunkel, 1994). To obtain PCR products containing the smallest possible number of introduced errors, a two stage PCR process was developed which employed a *Taq* polymerase for the initial rounds of amplification to maximise sensitivity, with a proof-reading mixture of enzymes (Expand™, Boehringer-Mannheim) used for the remainder of the amplification to minimise the introduction of PCR errors. Figure 4.9 illustrates the optimisation of template concentration for the stage II PCR and shows that yields obtained using Expand™ proofreading enzyme mixture for stage II PCR are comparable to those obtained using *Taq* polymerase. Direct sequencing of PCR products, while being a relatively rapid process, could lead to sequence ambiguities if errors are introduced during early rounds of amplification (Sogin, 1990; Wrischnik *et al.* 1987). Potential problems of this type could be avoided by mixing the products of multiple PCR reactions, which should reduce the

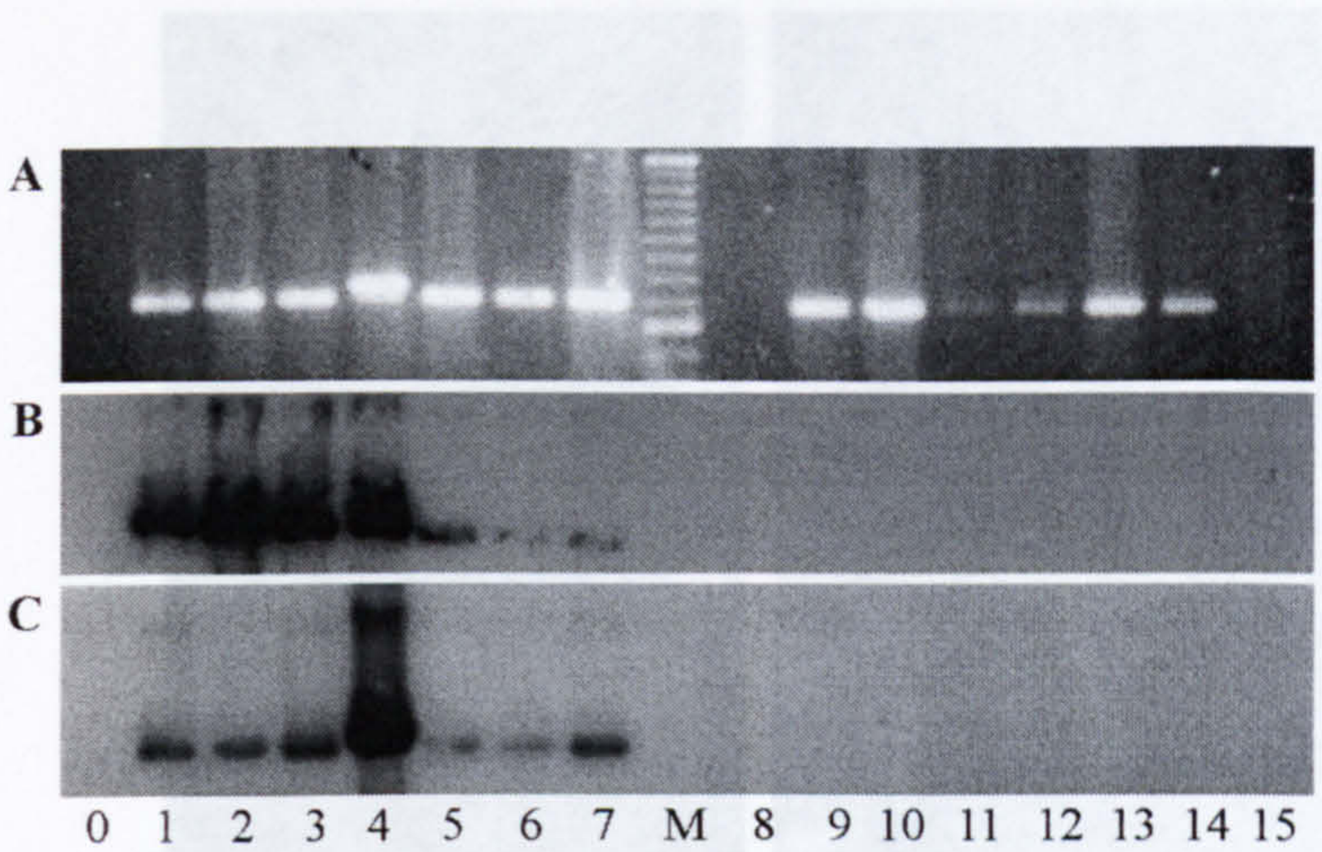


Figure 4.4 ***Aspergillus parasiticus*** **major and minor LC1/2c PCR products hybridise differentially to other LC1/2c PCR products.**

A. PCR products obtained using genomic template DNA from a range of deuteromycete fungi and two primer-pairs, LC1/2c (lanes 0-7) and LC3/5c (lanes 8-15). See legend to figure 4.3(A) for reaction conditions and template identities. These PCR products were blotted from duplicate gels onto nylon membranes and probed with ^{32}P-labelled *A. parasiticus* LC1/2c PCR products.

B. Probed with minor product.

C. Probed with major product.

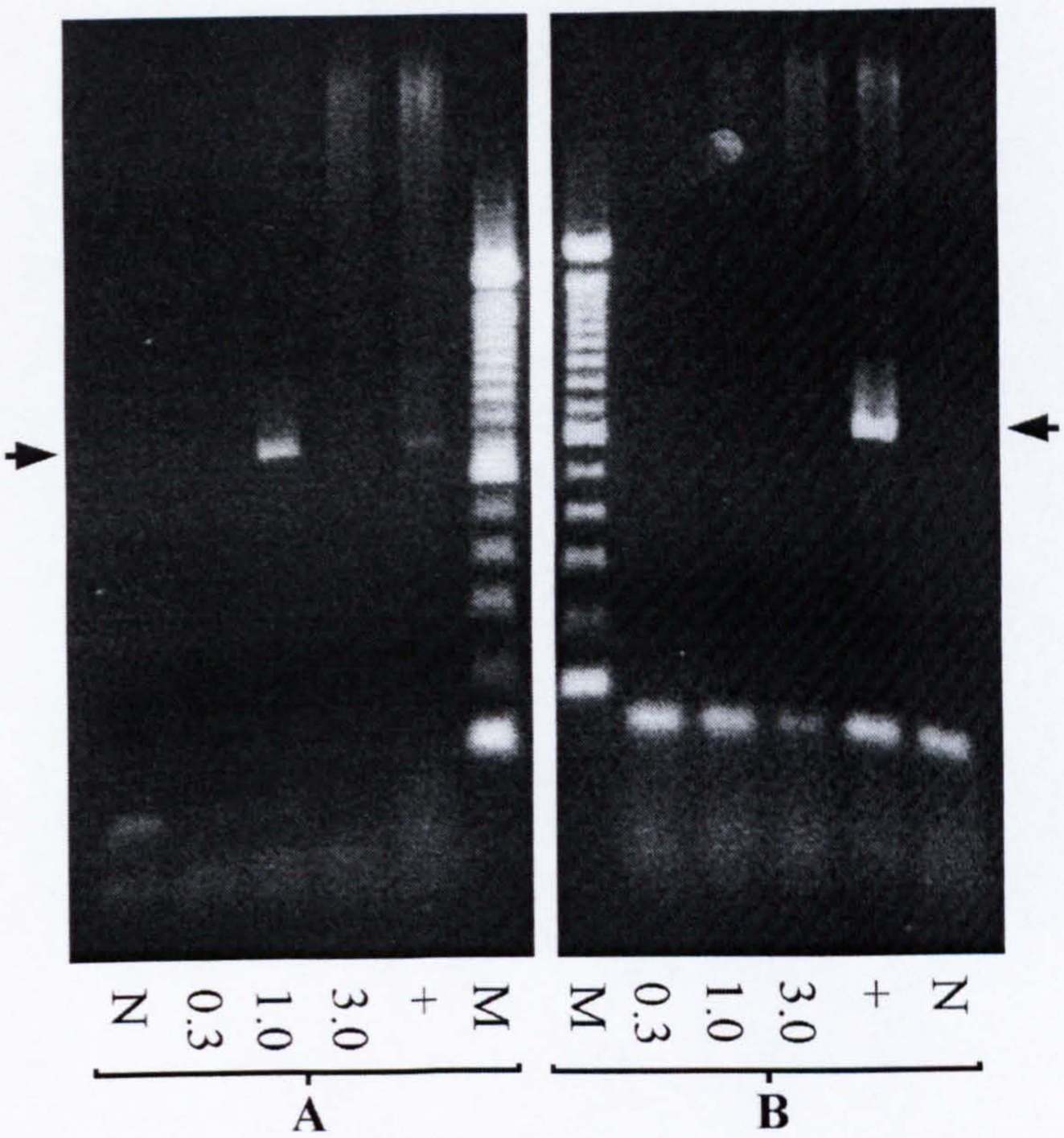


Figure 4.5 A PCR product is amplified from *Verticillium dahliae* genomic DNA using the LC1/2c primer pair.

PCR was carried out on three different amounts (0.3, 1.0 and 3.0 µl DNA/25 µl reaction) of *V. dahliae* genomic DNA solution (100 ng/µl), a 10^{-1} dilution of the original DNA preparation, using LC1/2c (**A**) and LC3/5c (**B**) primer pairs. Thermal cycling parameters for all reactions were: (94° C, 3') x 1, (94° C, 1'; 55° C, 1'; 72° C, 3') x 40, (72° C, 10') x 1. PCR primers were used at the following concentrations: LC1, 3.2 µM; LC2c, 0.4 µM; LC3 1.2 µM; LC5c, 0.8 µM. Reaction products were fractionated on a 1.2% agarose gel in 0.5 x TBE buffer at 80 mA. A PCR product of approximately 750 bp (arrowed) was obtained using the LC1/2c primer pair with 1 µl of template DNA solution. **M**, 100 bp ladder DNA size markers (Gibco-BRL); **N**, no DNA (negative control); +, *P. patulum* genomic DNA (positive control).

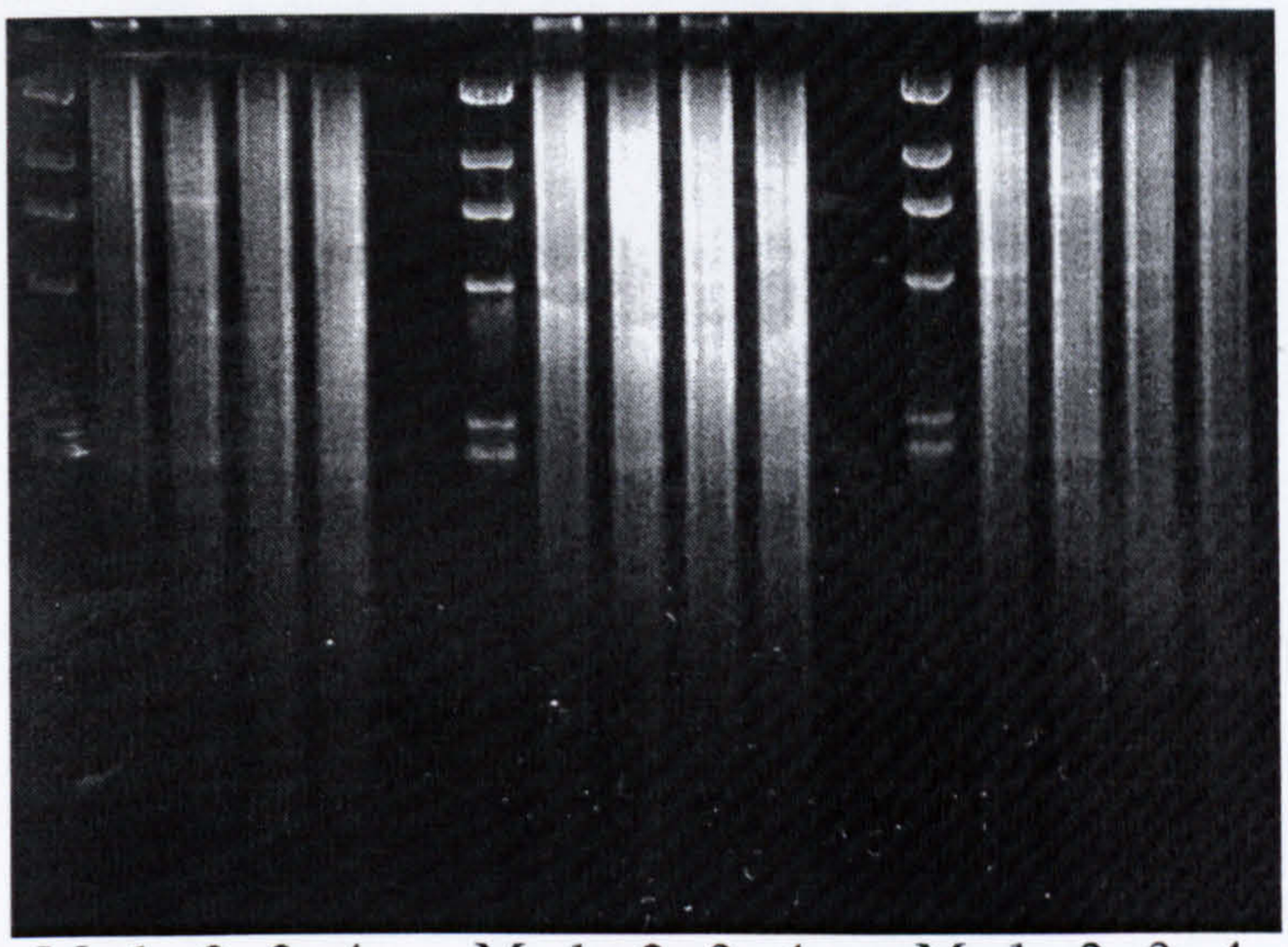


Figure 4.6 Restriction digests of genomic DNA for Southern blotting.

Genomic DNA from *Penicillium patulum* (**1**, **2**)and *Aspergillus parasiticus* (**3**, **4**)was digested overnight at 37° C with restriction enzymes *Bam*HI (**1**, **3**) and *Eco*RI (**2**, **4**). Approximately 5 μg of DNA from each digest was fractionated on a 1% agarose gel in TAE buffer at 40 mA/40 V. **M**, λ *Hin*dIII DNA size markers.

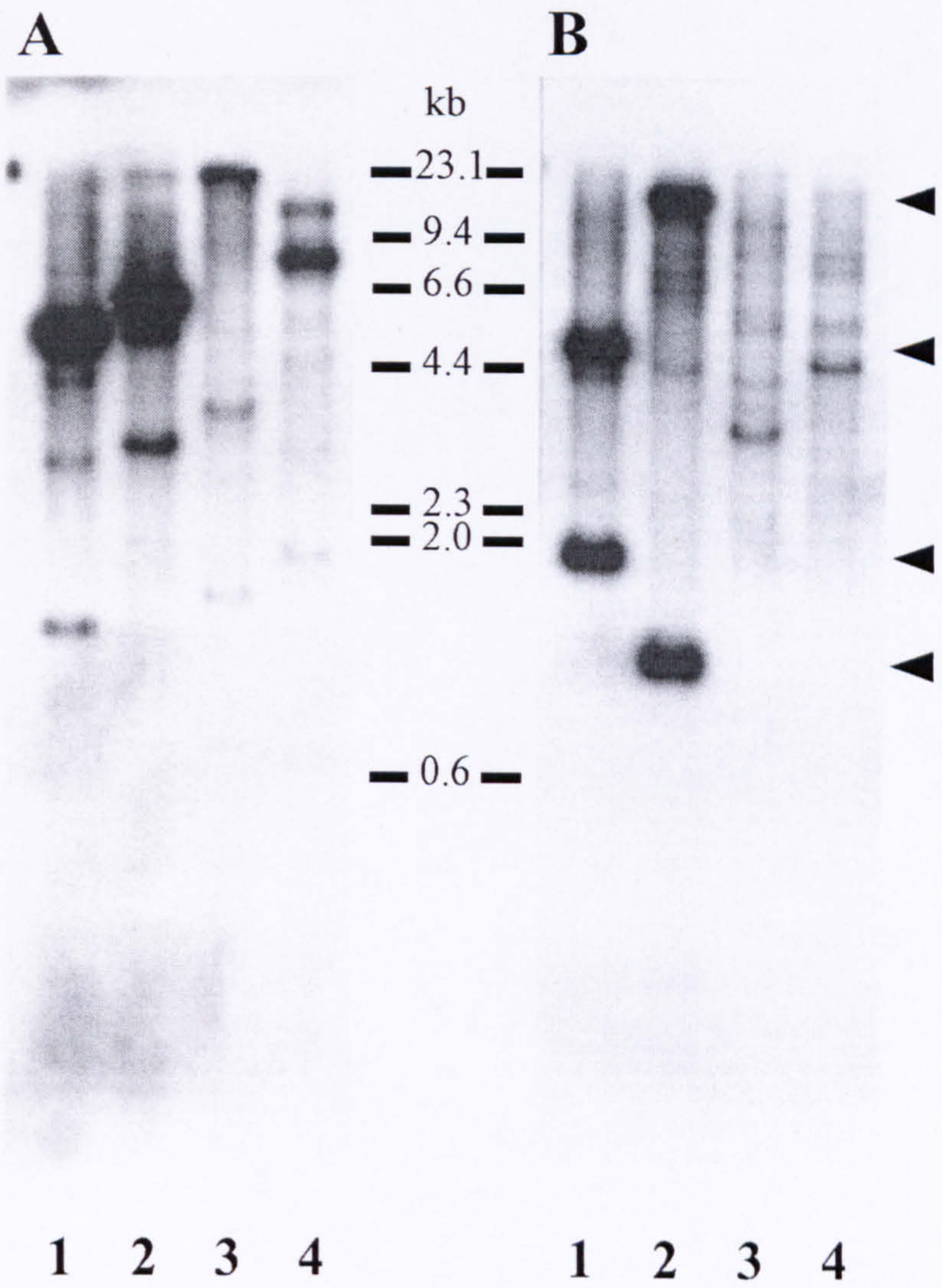


Figure 4.7 Genomic Southern blots probed with *Penicillium patulum* LC1/2c and LC3/5c PCR products.

Restriction digests shown in figure 4.6 were blotted onto nylon membrane and probed with ^{32}P-labelled *P. patulum* LC1/2c and LC3/5c PCR products. Probes were hybridised to the membrane and washed at 55° C. **1**, *P. patulum* DNA digested with *Bam*HI; **2**, *P. patulum* DNA digested with *Eco*RI; **3**, *A. parasiticus* DNA digested with *Bam*HI; **4**, *A. parasiticus* DNA digested with *Eco*RI. **A**, probed with LC1/2c product; **B**, probed with LC3/5c product. The positions of λ *Hin*dIII DNA size markers are indicated between the panels. Hybridisations in panel B, lanes 1 and 2, corresponding to restriction fragments of *MSAS* (Beck *et al.*, 1990) and *pks2* (chapter 5) are arrowed.

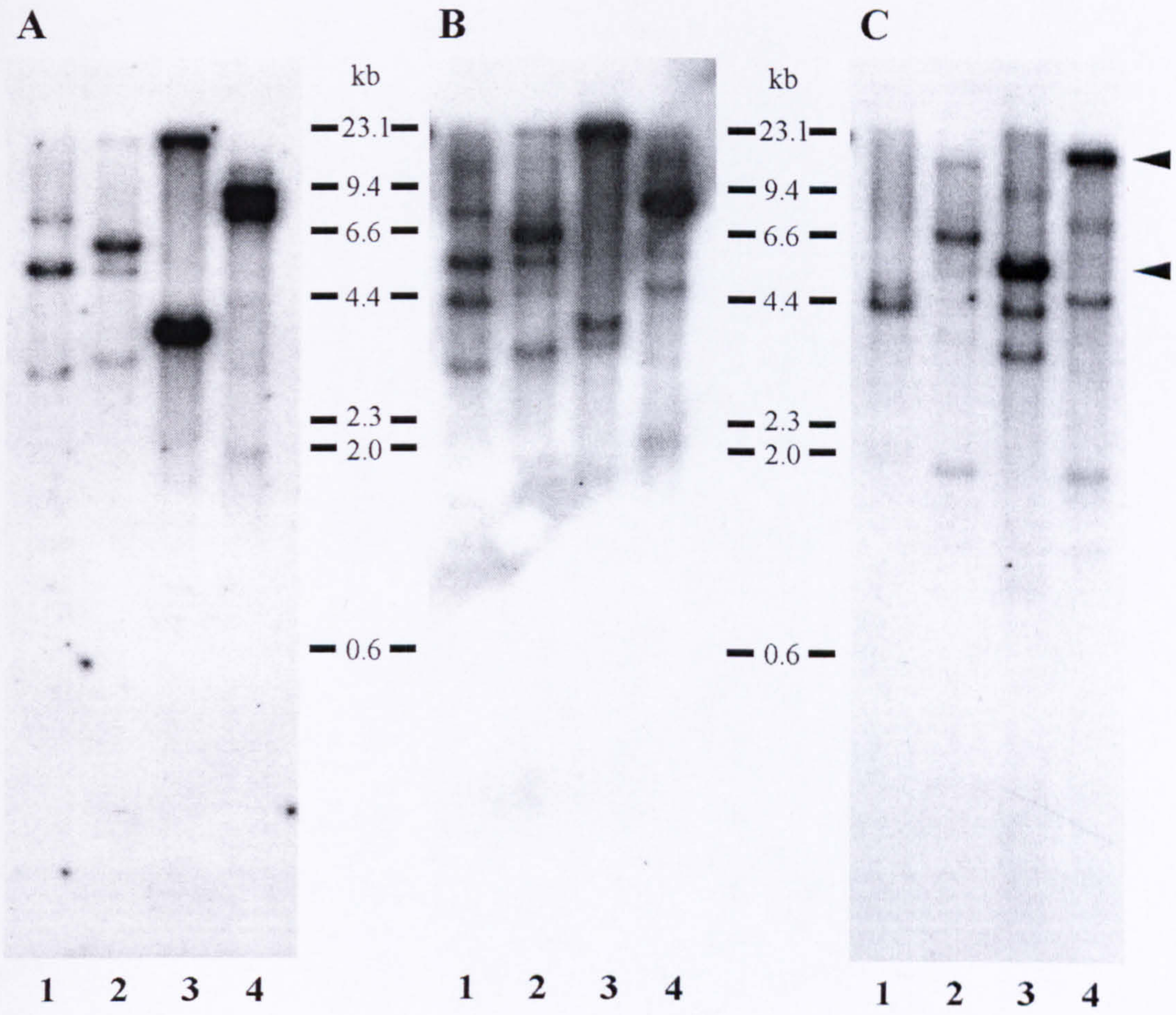


Figure 4.8 Genomic Southern blots probed with *Aspergillus parasiticus* LC1/2c and LC3/5c PCR products.

Restriction digests shown in figure 4.6 were blotted onto nylon membrane and probed with ^{32}P-labelled *A. parasiticus* LC1/2c and LC3/5c PCR products. Probes were hybridised to the membrane and washed at 55° C. **1,** *P. patulum* DNA digested with *Bam*HI; **2**, *P. patulum* DNA digested with *Eco*RI; **3**, *A. parasiticus* DNA digested with *Bam*HI; **4**, *A. parasiticus* DNA digested with *Eco*RI. Blot **A**, probed with LC1/2c major product; blot **B**, probed with LC1/2c minor product; blot **C**, probed with LC3/5c product. The positions of λ *Hin*dIII DNA size markers are indicated between the panels. Hybridisations in panel C, lanes 3 and 4, corresponding to restriction fragments of *APPKS1* (Walsh, 1993) are arrowed.

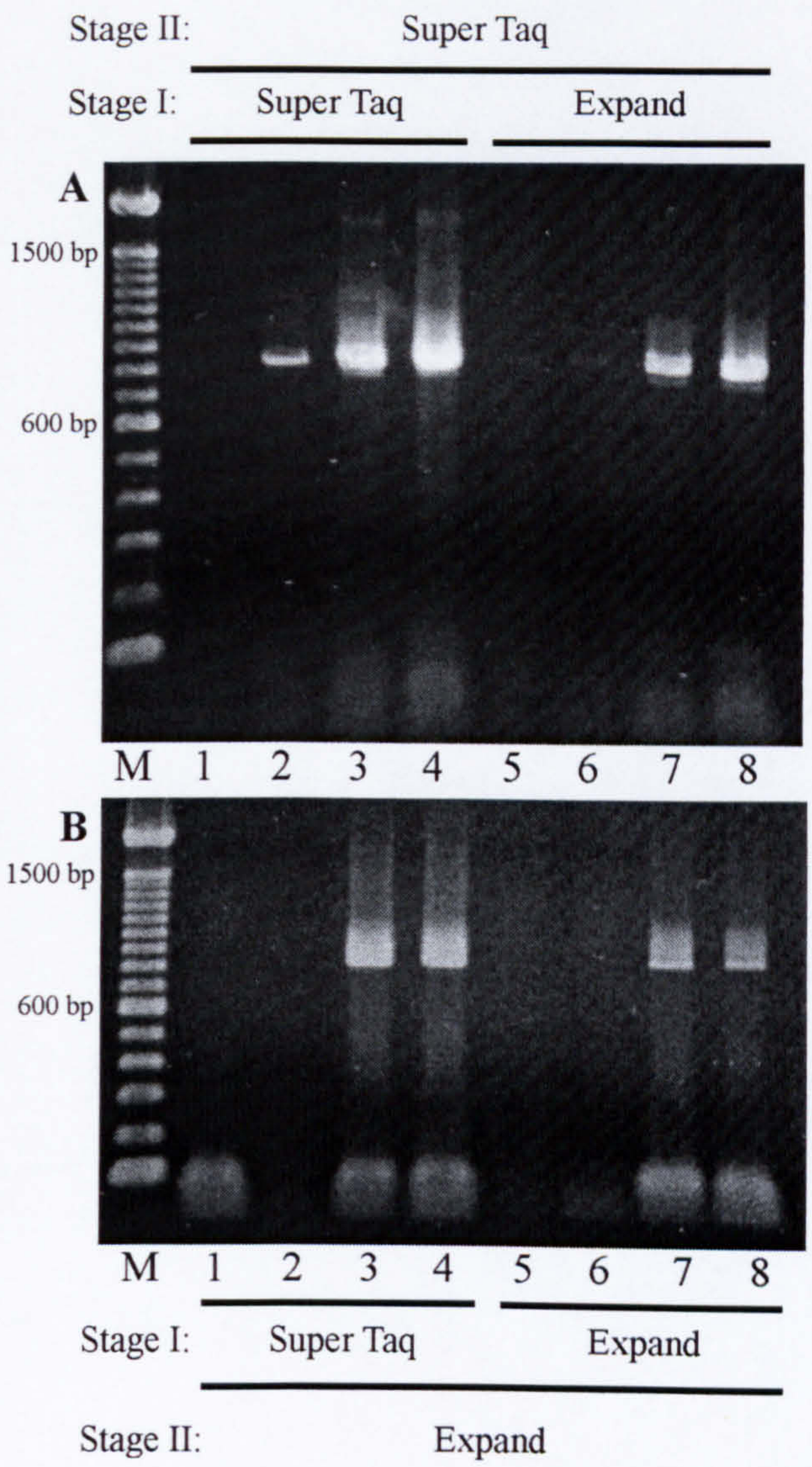


Figure 4.9 Optimising template concentration for two-stage PCR using LC-EXT primers with Super Taq and Expand™ DNA polymerases.

Reactions shown in **A** used Super Taq (HT Biotechnologies) for stage II PCR; Reactions shown in **B** used Expand™ (Boehringer-Mannheim) for stage II PCR. Reactions labelled **1-4** used Super Taq for stage I PCR, reactions labelled **5-8** used Expand™ for stage I PCR. **1** & **5,** negative controls. **M**, 100 bp ladder DNA size markers. Templates were: Stage I, 220 ng *A. parasiticus* genomic DNA or water (negative control); Stage II, 1-5 µl Stage I reaction products or 5 µl Stage I negative control reaction products. Thermal cycling parameters were: Stage I, Super Taq; (94° C, 3') x 1, (94° C, 1'; 55° C, 1'; 72° C, 3') x 5, (72° C, 6') x 1. Stage I, Expand™; (94° C, 2') x 1, (94° C, 15"; 55° C, 30"; 72° C, 1') x 10, (72° C, 6') x 1. Stage II, Super Taq; (94° C, 3') X 1, (94° C, 45"; 55° C, 45"; 72° C, 1'30") X 40, (72° C, 6') X 1. Stage II, Expand™; standard thermal cycling program (see chapter 2). Primer concentrations for all reactions were: LC1EXT, 3.2 µM; LC2cEXT, 0.4 µM. Reaction products were fractionated on a 1.2% agarose gel in 0.5 x TBE buffer at 55 mA.

effect of errors present in any one reaction. This approach was used to obtain DNA sequence for the LC1/2c PCR product from *V. dahliae.* Stage I and II PCR reactions (figure 4.9) were carried out in triplicate and the products pooled and sequenced using the STREP1T primer (figure 4.2). Although direct sequencing was successful in this instance, in general it could not be known whether products arose from a single template gene or from a mixture of genes. For this reason sequencing of cloned PCR products was felt to be a more generally useful approach. *Penicillium patulum* and *A. parasiticus* LC1/2c PCR products were digested at the *Bam*HI and *Sac*I restriction sites introduced by the LC-EXT tailed primers and cloned by ligation into the pUBS1 vector. Plasmid minipreps were obtained for five clones of each PCR product and a mixed template, consisting of equal amounts of DNA from each clone, was sent for automated DNA sequencing. The DNA sequences obtained are shown in figure 4.10, together with the predicted amino acid sequences they encode. A FASTA similarity search of the GenBank and EMBL databases indicated that each product was highly homologous to various fungal and bacterial type I PKS genes (figure 4.11). The DNA sequences were checked for the presence of *Bam*HI and *Eco*RI restriction sites; the only site identified was a single *Eco*RI site in the *A. parasiticus* major product (figure 4.10). This is consistent with the observed strong hybridisations of this product to two bands in a Southern blot of *Eco*RI-digested genomic DNA (figure 4.8A).

Figure 4.10 DNA and predicted amino acid sequences for PCR products obtained using LC1/2cEXT tailed degenerate primers.

The *Aspergillus parasiticus* and *Penicillium patulum* DNA sequences were obtained by sequencing of cloned PCR products. A gap (-) has been inserted into the *A. parasiticus* major product predicted amino acid sequence to maximise alignment with the other sequences. The *Eco*RI restriction site in the *A. parasiticus* major product DNA sequence is highlighted in italic type.

All sequences are shown in the direction LC1→LC2c and primer binding sites are highlighted in bold type. The predicted intron in the major *Aspergillus parasiticus* DNA sequence (section 4.4.2) is underlined. The *Verticillium dahliae* DNA sequence is the complementary strand to that obtained by direct sequencing of the PCR product using the STREP1T primer and is incomplete. The predicted amino acid sequence for this product has been numbered in alignment with the other PCR products

Aspergillus parasiticus major LC1/2cEXT PCR product.

```
  1  GATCCGAGGT TCTTCAACAT GTCCCGAAAG AACTCTTCAA AAGATCCGAT
 51  GCAGCGATTG GCGCTTGTCA CCGCTTACGA GGCACTGGAG ATGTCTGGAT
101  ATGTTGGTAA TAGGACACGC TCGACCACTC TAAGTCGTAT TGGGACTTTC
151  TACGGCCAGA CAAGTGATGA TTATCGGGAT GTCAACGCTG CACAGGATAT
201  TGGGACGTAT TTCATCACTG GTGGAATTCG TGCATTTGGA CCTGTAGGTG
251  TTGATGTATA CCCTTTACTA GCGTATCACA TTGCCAACAA TTTTACAGGG
301  ACGAATCAAT TACTACTTCA AATTCGAAGG TCCCAGTTTT AGTGTAGATA
351  CTGCCTGTTC CTCTAGCCTG GCTGCGATCC AGCTAGCCTG TACATCACTG
401  TGGAGTGGTG ATTGCGACAC GGCAGTTACG GGTGGACTCA GCGTACTCAC
451  GTCGCCTGAT CTCTTTTCAG GATTGAGTCG AGGCCAGTTT CTTTCACAGA
501  CTGGGTCATG CAAGACCTTT GACAAAGCTG CTGATGGCAA TTGCCGCGCA
551  GATGGAGTGG GAACGGTGGT TATCAAAAGG CTAAAAGATG CTGAGCTTGA
601  CAATGATAAT GTTCTGGCTG TAATTCTCGG TGCTGCCACA AATCATTCAG
651  CGCAGGCAGT CAGCATTACG CATCCACATG CAGAGACCCA ATCCAAGCTA
701  TATCGCGAGA TCTTGCAACA ATCAGGTGTC GATCCCTTTG ATGTGGGCTA
751  TGTCGAAATG CACGGCACSG GCAC
```

Predicted polypeptide encoded by *A. parasiticus* major PCR product.

```
  1  DPRFFNMS-R KNSSKDPMQR LALVTAYEAL EMSGYVGNRT RSTTLSRIGT
 51  FYGQTSDDYR DVNAAQDIGT YFITGGIRAF GPGRINYYFK FEGPSFSVDT
101  ACSSSLAAIQ LACTSLWSGD CDTAVTGGLS VLTSPDLFSG LSRGQFLSQT
151  GSCKTFDKAA DGNCRADGVG TVVIKRLKDA ELDNDNVLAV ILGAATNHSA
201  QAVSITHPHA ETQSKLYREI LQQSGVDPFD VGYVEMHGTG
```

Aspergillus parasiticus minor LC1/2cEXT PCR product.

```
  1  GACCCGAGGT TCTTCAACAT GTCGCCACGC GAAGCCCTCC AGGCAGATCC
 51  CGCTCAAAGA CTTGCATTGC TCACGGCCTA TGAAGCTCTT GAAATGGCCG
101  GCTTTATCCC CGACAGCACC CCTTCTACCC AGAGGGATCG AGTTGGCCTT
151  TTCTATGGAA TGACTAGCGA TGACTATCGT GAGATAAATA GTGGTCAAGA
201  TATTGATACA TACTTTATCC CTGGTGGGAA TCGTGCTTTC ACACCTGGTC
251  GGATAAACTA CTATTTCAAG TTCAGTGGGC CCAGTGTCAG CGTTGATACA
301  GCTTGTTCTT CAAGTCTTGC GGCTATTCAT ATGGCTTGCA ATTCGATCTG
351  GAGAAATGAT TGCGATGCTG CTATTGCTGG AGGTGTCAAT ATATTGACAA
401  ACCCTGATAA CCATGCCGGT CTCGACCGTG GCCATTTCCT GTCCAGAACC
451  GCGGAATTGTA ACACATTTGA CGATGGTGCT GATGGCTACT GTAGAGCAGA
501  CGGAGTGGGT ACGATTATCC TCAAGCGGCT GGAAGATGCT CAGGTGGACA
551  ACGATCCAAT CCTCGGCGTG ATCAATGGGG CCTATACCAA CCATTCGGCA
601  GAAGCAGTCT CGATTACCCG CCCTCATGTT GGCGCACAAG CGTTTATCTT
651  TAATAAACTA TTGAACGATG CCAATATCGA CCCAAAGGAC GTCAGCTACG
701  TTGAGATGCA CGGCACCGGC AC
```

Predicted polypeptide encoded by *A. parasiticus* minor PCR product.

```
  1  DPRFFNMSPR EALQADPAQR LALLTAYEAL EMAGFIPDST PSTQRDRVGL
 51  FYGMTSDDYR EINSGQDIDT YFIPGGNRAF TPGRINYYFK FSGPSVSVDT
101  ACSSSLAAIH MACNSIWRND CDAAIAGGVN ILTNPDNHAG LDRGHFLSRT
151  GNCNTFDDGA DGYCRADGVG TIILKRLEDA QVDNDPILGV INGAYTNHSA
201  EAVSITRPHV GAQAFIFNKL LNDANIDPKD VSYVEMHGTG
```

Penicillium patulum LC1/2cEXT PCR product.

```
  1  GATCCGAGGT TCTTCAACAT GTCTCCCCGC GAAGCCCTTC AAGCAGACCC
 51  TGCTCAGCGT CTGGCACTTC TTACTGCTTA CGAGGCTTTG GAGATGGCAG
101  GCTTTATCCC TGATAGCACT CCCTCCACGC AGAAAAATCG CGTGGGTGTC
151  TTCTATGGAA TGACCAGTGA TGACTACCGT GAGGTCAATA GTGGGCAGGA
201  TATTGATACC TACTTCATCC CTGGTGGTAA CCGTGCTTTC ACTCCCGGTC
251  GCATCAATTA CTATTTCAAG TTCAGCGGTC CAAGTGTCAG CGTGGATACA
301  GCTTGCTCGT CAAGTCTTGC GGCAATTCAT GTCGCCTGTA ATTCCTTGTG
351  GAGGAATGAA AGTGATTCTG CCGTGGCTGG CGGCGTCAAT ATCCTGACAA
401  ATCCTGACAA CCATGCTGGT CTTGATCGTG GACATTTCCT GTCAAGGACT
451  GGAAACTGCA CCACATTTGA TGATGGTGCT GATGGTTATT GTAGAGCAGA
501  TGGAATTGGC TCCGTTGTGT TGAAGAGACT CGAGGATGCT CAGGCAGACA
551  ACGATCCGAT TTATGGTATC ATTGGAGGAG CCTACACCAA CCACTCAGCG
601  GAGGCTGTGT CAATTACTCG GCCTCACGTA GGGGCTCAGT CTTTTATCTT
651  TGACAAGCTT CTCAACGAGT CCAACAGCGA TCCAAAAGAG GTCAGCTACA
701  TCGAGATGCA TGGCACCGGC AC
```

Predicted polypeptide encoded by *P. patulum* PCR product.

```
  1  DPRFFNMSPR EALQADPAQR LALLTAYEAL EMAGFIPDST PSTQKNRVGV
 51  FYGMTSDDYR EVNSGQDIDT YFIPGGNRAF TPGRINYYFK FSGPSVSVDT
101  ACSSSLAAIH VACNSLWRNE SDSAVAGGVN ILTNPDNHAG LDRGHFLSRT
151  GNCTTFDDGA DGYCRADGIG SVVLKRLEDA QADNDPIYGI IGGAYTNHSA
201  EAVSITRPHV GAQSFIFDKL LNESNSDPKE VSYIEMHGTG
```

***Verticillium dahliae* LC1/2cEXT PCR product.**

```
  1  ACATTTTACG TTCAGACTTC GGATGATTGG CGAGAGATCA ATGCCGCACA
 51  GGAAGTGGAA ACATATTTCA TCACGGGGGG AGTACGCGCC TTTGGCCCTG
101  GCAGAATCAA CTACCACTTC GGTTTCAGTG GACTTAGTcT GAACATCGAT
151  ACAGCCTGTT CTTCCAGTGC CGCAGCTCTG CAAATCGCAT GCACTTCGcT
201  cTGGGCCAAG GACTGTGACA CGGCAGTCGT TGGCGGCTTG TCTTGCATGA
251  CCAACCCCGA TATTTTCTCA GGACTCAGCC GTGGCCAGTT CTTGTCCAAG
301  ACTGGTCCCT GTGCTACGTT TGACAATGGA GCTGATGGTT ACTGCCGAGC
351  TGATGGCTGC GCCTCCGTCA TTGTGAAGCG TCTGGATGAT GCCATTGCCG
401  ACAAGGACAA TGTCCTGGCG GTCATCCTGG GCACGGCTAC CAACCATTCT
451  GCCGACGCCA TATCGATTAC CCATCCTCAT GGACCTACGC AGTCGATCCT
501  ATCATCAGCC ATTCTCGATG AGGCTGGCGT CGTCCTCTCG ATGTTGATAC
551  GTCGAGAGCA C
```

Predicted polypeptide encoded by *V. dahliae* PCR product.

```
 50                                                        T
 51  FYVQTSDDWR EINAAQEVET YFITGGVRAF GPGRINYHFG FSGLSLNIDT
101  ACSSSAAALQ IACTSLWAKD CDTAVVGGLS CMTNPDIFSG LSRGQFLSKT
151  GPCATFDNGA DGYCRADGCA SVIVKRLDDA IADKDNVLAV ILGTATNHSA
201  DAISITHPHG PTQSILSSAI LDEAGVVLSM LIRREH
```

Figure 4.11 Results of similarity searches of GenBank database using the LC1/2c PCR products as query sequences.

The GenBank database was searched for sequences similar to the LC1/2c PCR products from *A. parasiticus*, *P. patulum* and *V. dahliae*, using the FASTA program (Wisconsin package version 9.0, Genetics Computer Group) using a gap creation penalty of 16 and a gap extension penalty of 4 The five sequences with the highest z-scores (z-sc) , i.e. the most similar to the query sequence, are shown for each LC1/2c PCR product. Multiple database entries for a single sequence have been removed. The E( ) score represents the number of sequences expected to produce a z-score by chance equal or greater to that obtained in the search. Genbank sequence ID and accession number for each sequence is shown, along with the locations of the beginning and end of homology to the query sequence.

A. parasticus minor product

```
The five best scores are:                                  z-sc   E(588376)
1. Gb_Pl:Anwa      Begin: 3011  End:  3732
X65866 A.nidulans wA                                       2191.2       0
2. Gb_Pl:Cogpks1   Begin: 1852  End:  2573
D83643 C.lagenarium PKS1                                   1321.9       0
3. Gb_Pl:Asnpks1   Begin: 1875  End:  2596
L42765 A.parasiticus pksL1                                 1175.1       0
4. Gb_Pl:Appksagen Begin: 2897  End:  3618
Z47198 A.parasiticus pksA                                  1174.8       0
5. Gb_Pl:Emestca   Begin: 2211  End:  2932
L39121 A.nidulans pksST...                                 1133.6       0
```

A. parasiticus major product

```
The five best scores are:                                  z-sc   E(589198)
1. Gb_Pl:Anwa      Begin: 3247  End:  3732
X65866 A.nidulans wA                                        859.9       0
2. Gb_Pl:Asnpks1   Begin: 2113  End:  2596
L42765 A.parasiticus pksL1                                  776.0       0
3. Gb_Pl:Appksagen Begin: 3135  End:  3618
Z47198 A.parasiticus pksA                                   775.7       0
4. Gb_Pl:Emestca   Begin: 2457  End:  2932
L39121 A.nidulans pksST                                     734.5   3.9e-34
5. Gb_Pl:Cogpks1   Begin: 1852  End:  2098
D83643 C.lagenarium PKS1                                    541.0   2.8e-23
```

P. patulum

```
The five best scores are:                                  z-sc   E(589198)
1. Gb_Pl:Anwa      Begin: 3011  End:  3732
X65866 A.nidulans wA                                       2199.9       0
2. Gb_Pl:Cogpks1   Begin: 1852  End:  2573
D83643 C.lagenarium PKS1                                   1375.6       0
3. Gb_Pl:Asnpks1   Begin: 1875  End:  2596
L42765 A.parasiticus pksL1                                 1232.4       0
4. Gb_Pl:Appksagen Begin: 2897  End:  3618
Z47198 A.parasiticus pksA                                  1232.1       0
5. Gb_Pl:Emestca   Begin: 2211  End:  2932
L39121 A.nidulans pksST                                    1164.7       0
```

V. dahliae

```
The five best scores are:                                  z-sc   E(584793)
1. Gb_Pl:Cogpks1   Begin: 1999  End:  2562
D83643 C.lagenarium PKS1                                   1697.5       0
2. Gb_Pl:Anwa      Begin: 3161  End:  3716
X65866 A.nidulans wA                                        805.4       0
3. Gb_Pl:Appksagen Begin: 3066  End:  3602
Z47198 A.parasiticus pksA                                   584.5    1e-25
4. Gb_Pl:Asnpks1   Begin: 2044  End:  2580
L42765 A. parasiticus pksL1                                 584.8   1.1e-25
5. Gb_Pl:Emestca   Begin: 2380  End:  2849
L39121 A.nidulans pksST                                     543.9   1.6e-23
```

4.3 DISCUSSION

4.3.1 The LC-series primers amplify fragments from two structurally distinct groups of PKS ketosynthase domains.

Using two pairs of degenerate primers, designed to amplify fragments containing PKS gene condensing domains, products of similar size were obtained from a range of deuteromycete fungi (figure 4.3). Hybridisations between PCR products from different genomes showed that the set of products from each primer-pair was structurally related, while the two sets of products were structurally distinct (figures 4.3 and 4.5). This pattern of hybridisation may indicate two subclasses of PKS genes, both widely distributed amongst the filamentous fungi. Cross-hybridisation between the two sets of PCR products is minimal in Southern blotting of both genomic DNA and PCR products (figures 4.3, 4.5, 4.7 and 4.8), suggesting that heterologous probing for genes in one subgroup with a gene from the other subgroup would not be effective. The PCR primers presented here may facilitate the cloning of a range of novel fungal PKS genes by providing highly homologous probes for library screening.

When used to probe genomic southern blots, these PCR products hybridised to multiple restriction fragments even at fairly high stringency (figures 4.7 and 4.8). This may indicate the presence of multiple PKS genes in each fungal genome. As some genera of filamentous fungi are known to produce several polyketide metabolites, requiring more than one class of polyketide backbone, the presence of multiple PKS genes in a single strain would be expected.

4.3.2 The major LC1/2c PCR product from *A. parasiticus* contains a putative intron.

The fact that the size of the PCR products obtained using these primers was so well conserved across a range of fungal genera suggested that the LC1/2c primers may have been amplifying across an intron in the *A. parasiticus* genome to produce the major PCR product from this species, which is approximately 50 bp larger than the other products. By comparing regions of continuous homology of predicted amino acid sequence between this PCR product and the other LC1/2c PCR products, a putative intron of 55 bp was identified within the *A. parasiticus* major product (figure 4.10). This putative intron possesses perfect 5' and 3' splice site consensus sequences of GTAGGT and CAG, although the nearest

match to the consensus internal element sequence (NNCTRAY) is TACTAGC at 26 bp from the 3' end (see section 3.3.1 for a description of consensus sequences). The proposed position of this intron is not conserved in any other fungal PKS sequence published to date.

4.3.3 Alignment of predicted protein sequences encoded by the sequenced LC1/2c PCR products reveals a high degree of conservation with other PKS ketosynthase domains.

Available fungal PKS condensing domain sequences were aligned with the predicted protein sequence for the LC1/2c PCR products and with non-fungal PKS and FAS condensing domains (figure 4.12). Sequence similarity between the fungal FAS and PKS condensing domains was so low it was not possible to include a fungal FAS in the alignment using the software available. The similarity between the fungal PKS and the mammalian FAS is much closer, suggesting that the fungal PKS genes have not evolved from the fungal FAS, but possibly from the FAS of some ancestral organism. To quote Vining (1992), "...the concept of fatty acid synthase gene duplication and subsequent evolution giving rise to a polyketide synthase in each individual organism appears untenable". The condensing domains of the type I PKS and FAS proteins are conserved throughout most of the region examined, however the extent of conservation with the type II condensing domain from *Streptomyces coelicolor* (actI) is much lower.

The region around the active site cysteine identified as binding the acetyl CoA substrate in the goose FAS (Poulose *et al.*,1984) is absolutely conserved within each of the predicted two subgroups of fungal PKS. Interestingly though, the MSAS-like PKSs are divergent from the other fungal and non-fungal PKSs and from the mammalian FAS examined in this region. All the sequenced LC1/2c PCR products and other genes in the *wA*-type subgroup exhibit the motif DTACSSSL (except for *C. lagenarium PKS1*, which diverges at the final residue) which is absolutely conserved with the mammalian FAS and the *eryA* ketosynthase active sites, while the *MSAS*-type genes possess the active site motif DAACASSL.

Figure 4.12 Alignment of cloned LC1/2c PCR products with other PKS and FAS ketosynthase domains.

Various ketosynthase regions were aligned to LC1/2c PCR product sequences using the PILEUP program (Wisconsin Package, Devereux *et al.*, 1984). Sequence conservation of 80% or greater (including conservative substitutions) in a column is indicated by shading of the conserved residues. The active site cysteine is indicated by an arrow. Sequence identification codes are: *Colletotrichum lagenarium* PKS1, Clpks1ks; *Aspergillus parasiticus* major PCR product, Lcex_apmaj; *A. nidulans* wA, Anwa; *A. parasiticus* minor PCR product, Lcex_Apmin; *Penicillium patulum* MSAS, Lcex_Ppks; *A. nidulans* pksST, Emestcaks; *A. parasiticus* pksL1, Pksl1ks; *A. terreus* MSAS, At6msasks; *P patulum* PKS2, Pppks2ks; *Cochliobolus heterostrophus* PKS1, Chpks1ks; *Saccharopolyspora erythraea* eryA, Ery2cks; Rat FAS, Ratfasks; *Streptomyces coelicolor* ActI-Orf1, Actiks.

```
Clpks1ks   :   DPRFFNMSPREAFQTDPMQRMALTTAYEALEMCGYVPNRTPSTRLDRI--GTFYGQT   :    55
Lcex_Apmaj:   DPRFFNMS-RKNSSKDPMQRLALVTAYEALEMSGYVGNRTRSTTLSRI--GTFYGQT   :    54
Anwaks     :   DPRFFNMSPREALQADPAQRLALLTAYEALEGAGFVPDSTPSTQRDRV--GIFYGMT   :    55
Lcex_Apmin:   DPRFFNMSPREALQADPAQRLALLTAYEALEMAGFIPDSTPSTQRDRV--GLFYGMT   :    55
Lcex_Ppks  :   DPRFFNMSPREALQADPAQRLALLTAYEALEMAGFIPDSTPSTQKNRV--GVFYGMT   :    55
Emestcaks  :   DPRFFSISPKEAPQMDPAQRMALMSTYEAMERGGIVPDTTPSTQRNRI--GVFHGVT   :    55
Pksl1ks    :   DPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRDRI--GVFHGVT   :    55
At6msasks  :   DAAFFGVSPKEAEQMDPQQRLSLEVTWEALEDAGIPPQSLSGSE---T--AVFMGVN   :    52
Pp6msasks  :   DCQFFGISPKEAEQMDPQQRVSLEVASEALEDAGIPAKSLSGSD---T--AVFWGVN   :    52
Appks1ks   :   DASFFNISPKEAEQMDPQQRLALEVTWEALENAGIPLSSLSGSD---A--AVFMGVN   :    52
Pppks2ks   :   DAAFFAISPREAEQMDPQQRIALEVAWEALENAGISPSRLAGSD---T--SVYMGVN   :    52
Chpks1ks   :   DAPFFNVSPAEAAALDPQQRMLLECSYEAFENSGTPMSKIVGTD---T--SVFVSSF   :    52
Ery2cks    :   DAEFFGVSPREAAAMDPQQRLLLETSWELVENAGIDPHSLRGTA---T--GVFLGVA   :    52
Ratfasks   :   DASFFGVHPKQAHTMDPQLRLLLEVSYEAIVDGGINPASLRGTN---T--GVWVGVS   :    52
Actiks     :   DPVAEGFGPRELDRMDRASQFAVACAREAFAASGLDPDTLDPARVGVSLGSAVAAAT   :    57

Clpks1ks   :   SDDWREI--NAAQ---EVDTYYITGGVRAF------GPGRINYHFGFSGPSLNVDTAC   :   102
Lcex_Apmaj:   SDDYRDV--NAAQ---DIGTYFITGGIRAF------GPGRINYYFKFEGPSFSVDTAC   :   101
Anwaks     :   SDDYREV--NSGQ---DIDTYFIPGGNRAF------TPGRINYYFKFSGPSVSVDTAC   :   102
Lcex_Apmin:   SDDYREI--NSGQ---DIDTYFIPGGNRAF------TPGRINYYFKFSGPSVSVDTAC   :   102
Lcex_Ppks  :   SDDYREV--NSGQ---DIDTYFIPGGNRAF------TPGRINYYFKFSGPSVSVDTAC   :   102
Emestcaks  :   SNDWMET--NTAQ---NIDTYFITGGNRGF------IPGRINFCFEFSGPSYSNDTAC   :   102
Pksl1ks    :   SNDWMET--NTAQ---NIDTYFITGGNRGF------IPGRINFCFEFAGPSYTNDTAC   :   102
At6msasks  :   SDDYSKLLLEDIP---NVEAWM--GIGTAY---CGVPNRISYHLNLMGPSTAVDAAC   :   101
Pp6msasks  :   SDDYSKLVLEDLP---NVEAWM--GIGTAY---CGVPNRISYHLNLMGPSTAVDAAC   :   101
Appks1ks   :   SDDYGKLLLEDLP---HVEPWM--GIGTAY---CGFANRISYHLNLMGPSTAVDAAC   :   101
Pppks2ks   :   SDDYGKLVLEDLT---GVGAHM--GVGTAY---CGIPSRISYLLDLMGPSVALDAAC   :   101
Chpks1ks   :   ATDYTDMLWRDPE---SVPMYQCTNSGFSR---SNLANRISYSFDLKGPSVLVDTAC   :   103
Ery2cks    :   KFGYGEDTAA-AE---DVEGYS--VTGVAP---AVASGRISYTMGLEGPSISVDTAC   :   100
Ratfasks   :   GSEASEALSRDPE---TLLGYSMVGCQRAM------MANRLSFFFDFKGPSIALDTAC   :   101
Actiks     :   SLEREYLLLSDSGRDWEVDAAWLSRHMFDYLVPSVMPAEVAWAVGAEGPVTMVSTGC   :   114

Clpks1ks   :   SSSAAALNVACNSLWQKDCDTAIVGGLSCMTNPDIFAGLSRGQFLSKTGP------CA   :   154
Lcex_Apmaj:   SSSLAAIQLACTSLWSGDCDTAVTGGLSVLTSPDLFSGLSRGQFLSQTGS------CK   :   153
Anwaks     :   SSSLAAIHLACNSIWRNDCDTAITGGVNILTNPDNHAGLDRGHFLSRTGN------CN   :   154
Lcex_Apmin:   SSSLAAIHMACNSIWRNDCDAAIAGGVNILTNPDNHAGLDRGHFLSRTGN------CN   :   154
Lcex_Ppks  :   SSSLAAIHVACNSLWRNESDSAVAGGVNILTNPDNHAGLDRGHFLSRTGN------CT   :   154
Emestcaks  :   SSSLAAIHLACNSLWRGDCDTAVAGGTNMIFTPDGHTGLDKGFFLSRTGN------CK   :   154
Pksl1ks    :   SSSLAAIHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDKGFFLSRTGN------CK   :   154
At6msasks  :   ASSLVAIHHGRQAILQGESEVAIVGGVNALCGPGLTRVLDKAGATSTEGR------CL   :   153
Pp6msasks  :   ASSLVAIHHGVQAIRLGESKVAIVGGVNALCGPGLTRVLDKAGAISSDGS------CK   :   153
Appks1ks   :   ASSLVAIHLGRQAILSGESKVAIVGGVNAIFGPGLTSVLDKAGALSSDGR------CH   :   153
Pppks2ks   :   ASSLVAVHHARQAIRAGETDLAIAGGVNALLGPGLTRVLDEAGAISADGK------CR   :   153
Chpks1ks   :   SGGLTALHLACQSLLVGDVRQALAAGSSLILGPEMMVTMSMMKFLSPDGR------CY   :   155
Ery2cks    :   SSSLVALHLAVESLRKGESSMAVVGGAAVMATPGVFVDFSRQRALAADGR------SK   :   152
Ratfasks   :   SSSLLALQNAYQAIRSGECPAAIVGGINLLLKPNTSVQFMKLGMLSPDGT------CR   :   153
Actiks     :   TSGLDSVGNAVRAIEEGSADVMFAGAADTPITPIVVACFDAIRATTARNDDPEHASR   :   171

Clpks1ks   :   TFDNGADGYCRADGCASVIVKRLDDALADKDNVLAVILGTATNHSA-DAISITHPHG   :   210
Lcex_Apmaj:   TFDKAADGNCRADGVGTVVIKRLKDAELDNDNVLAVILGAATNHSA-QAVSITHPHA   :   209
Anwaks     :   TFDDGADGYCRADGVGTVVLKRLEDALADNDPILGVINGAYTNHSA-EAVSITRPHV   :   210
Lcex_Apmin:   TFDDGADGYCRADGVGTIILKRLEDAQVDNDPILGVINGAYTNHSA-EAVSITRPHV   :   210
Lcex_Ppks  :   TFDDGADGYCRADGIGSVVLKRLEDAQADNDPIYGIIGGAYTNHSA-EAVSITRPHV   :   210
Emestcaks  :   AFDDAADGYCRAEGVGTVFIKRLEDALAENDPILATILDIKTNHSA-MSDSMTRPFK   :   210
Pksl1ks    :   PYDDKADGYCRAEGVGTVFIKRLEDALADNDPILGVILDAKTNHSA-MSESMTRPHV   :   210
At6msasks  :   SFDEDAKGYGRGEGAAVVILKRLSTAIRDGDHIRAIIKGSAVAQDG-KTNGIMAPNA   :   209
Pp6msasks  :   SFDDDAHGYARGEGAGALVLKSLHRALLDHDNVLAVIKGSAVCQDG-KTNGIMAPNS   :   209
Appks1ks   :   SFDDTASGYGRGEGAAVVILKNMAEAVKNGDHILATLKGTAVAQDG-RTNGIMAPNQ   :   209
Pppks2ks   :   SFDDSANGYGRGEGAGVVILKRLEKALTDGDRVLAVLKGSAVASDG-KTLGIMAPNA   :   209
Chpks1ks   :   AFDERANGYARGEGVAVLLLKRLEDALADNDTIRAVIRGTGCNQDG-KTPGITMPNS   :   211
Ery2cks    :   AFGAGADGFGFSEGVTLVLLERLSEARRNGHEVLAVVRGSALNQDG-ASNGLSAPSG   :   208
Ratfasks   :   SFDDSGNGYCRAEAVVAVLLTKKSLARR----VYATILNAGTNTDGCKEQGVTFPSG   :   206
Actiks     :   PFDGTRDGFVLAEGAAMFVLEDYDSALARGARIHAEISGYATRCNAYHMTGLKA-DG   :   227

Clpks1ks   :   PTQSILSRAILDDAGVDPFLDVDYYVEMHGTG   :   240
Lcex_Apmaj:   ETQSKLYREILQQSGVDPFDVGYVEMHGTG   :   239
Anwaks     :   GAQAFIFKKLLNEANVDPKNISYIEMHGTG   :   240
Lcex_Apmin:   GAQAFIFNKLLNDANIDPKDVSYVEMHGTG   :   240
Lcex_Ppks  :   GAQSFIFDKLLNESNSDPKEVSYIEMHGTG   :   240
Emestcaks  :   PAQIDNMSALLSTAGISPLDLSYIEMHGTG   :   240
Pksl1ks    :   GAQIDNMTAALNTTGLHPNDFSYIEMHGTG   :   240
At6msasks  :   KAQELVAWNALRTAGVDPLTVGYVEAHATS   :   239
Pp6msasks  :   VAQQLAANNALSAANIDPHTVRYVEAHATS   :   239
Appks1ks   :   KAQELVARKALDVARVDASTIDYVEAHATS   :   239
Pppks2ks   :   QAQILVAQKALKEARVTPDSISYIEAHATS   :   239
Chpks1ks   :   VSQEALIRSVYKKAALDPLDTTYVECHGTG   :   241
Ery2cks    :   PAQRRVIRQALESCGLEPGDVDAVEAHGTG   :   238
Ratfasks   :   EAQEQLIRSLYQPGGVAPESLEYIEAHGTG   :   236
Actiks     :   REMAETIRVALDESRTDATDIDYINAHGSG   :   257
```

4.3.4 Phylogenetic analysis groups the LC1/2c PCR products with fungal PKS genes involved in spore pigment, melanin and aflatoxin biosynthetic pathways.

A cladogram has been constructed for the predicted proteins encoded by the LC1/2c PCR products, together with various other PKS and FAS genes (figure 4.13). The type II ActI ketosynthase from *Streptomyces coelicolor* has been included as an outgroup and used to root the tree. This diagram indicates possible evolutionary relationships between the different fungal PKS genes and the genes represented by the LC1/2c PCR products. As expected, the fungal PKS genes fall into two distinct clusters: the "*MSAS*-type" genes are in one cluster and the spore pigment, melanin and aflatoxin genes in the other cluster along with the LC1/2c PCR products. Genes in one cluster probably shared a common progenitor with each other more recently than with genes in the other cluster. A significant distinction between genes in the two clusters may be the presence or absence of a ketoreductase domain. The PKS and FAS genes clustering with *MSAS* all possess ketoreductase domains, while the identified PKS genes clustering with *A. nidulans wA* lack ketoreductase domains (figure 1.6), although this is as yet unconfirmed for the genes represented by the LC1/2c PCR products. The low bootstrap values for the grouping of non-fungal type II PKS and FASs within the "MSAS-type" gene cluster indicate that this grouping is probably not significant. Within the second cluster, the major *A. parasiticus* product is most closely related to the *C. lagenarium* melanin PKS, while the minor *A. parasiticus* and *P. patulum* products are more closely related to the *A. nidulans wA* gene product. The bootstrap values for these clusterings are high, suggesting that they reflect genuine evolutionary relationships. This arrangement suggests that the *A. nidulans wA* and the PKSs involved in biosynthesis of the octaketide aflatoxins and sterigmatocystin have evolved from a PKS involved in biosynthesis of the pentaketide 1,8-dihydroxynaphthalene (DHN) melanins. A similar evolutionary relationship between aflatoxins and octaketide spore pigment precursors in *Aspergillus*, based on structural similarities, has previously been proposed (Brown & Salvo, 1994). In the actinomycete *Streptomyces coelicolor,* a PKS gene involved in spore pigment biosynthesis has been suggested as a possible ancestor to the PKS genes responsible for antibiotic production, via the introduction of a promoter allowing expression in the substrate mycelium (Davis & Chater, 1990).

The predicted amino acid sequence for the *V. dahliae* PCR product (figure 4.10) was not included in the phylogenetic analysis shown in figure 4.13 because only a partial DNA

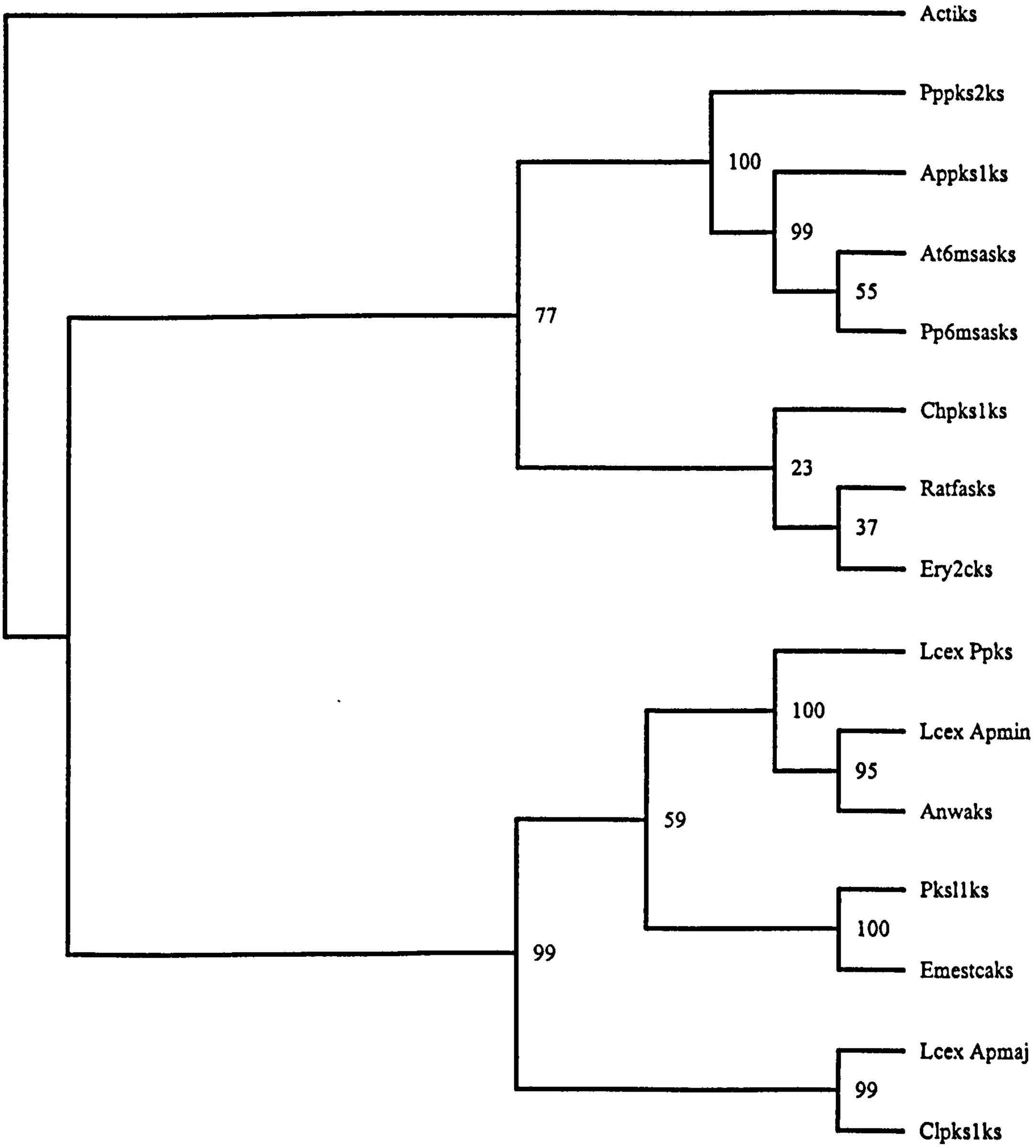


Figure 4.13 Phylogenetic relationships of LC1/2c PCR products and other PKS and FAS ketosynthase domains.

This tree was generated using the PHYLIP software package (J. Felsenstein, University of Washington). SEQBOOT was used to calculate 100 bootstrapped data sets from the sequence alignment shown in figure 4.12, which were then converted into a distance matrix by the program PROTDIST using Kimura's approximation to PAM distance. This protein distance matrix was converted into a set of trees by the program NEIGHBOR using the UPGMA method of clustering and from these the consensus tree shown was constructed using the program CONSENS. The resulting tree was displayed and printed using TREEVIEW (R.D.M. Page, University of Glasgow) with the ActI outgroup sequence used to root the tree. Bootstrap values are displayed at each node of the tree and relate to the group consisting of the species to the right of the node. For sequence identification codes see the legend to figure 4.12.

sequence had been obtained. This sequence is extremely close to that of the *C. lagenarium* melanin biosynthesis gene *PKS1* (as discussed in section 4.3.5) and would cluster with this gene if included in the analysis.

4.3.5 Speculation on the function of the putative PKS genes amplified by PCR using the LC-series primers.

Polyketides known to be produced by the fungal strains used in this study are listed in table 4.1. The PKS involved in the biosynthesis of any of these metabolites could be encoded by one of the putative PKS genes amplified with the LC-series degenerate primers.

The LC1/2c primer pair is likely to have primed on PKS genes involved in the biosynthesis of melanins or other spore pigments. The minor *A. parasiticus* and the *P. patulum* PCR products obtained using these primers show a very high degree of similarity to the *A. nidulans wA* gene (Mayorga & Timberlake, 1992) in both amino acid and DNA sequence, suggesting that they may be homologues of this gene. *wA* is required for the production of a yellow pigment intermediate in the biosynthesis of a dark green conidial pigment which confers resistance to ultraviolet light (Aramayo *et al.*, 1989). Green to blueish-green conidial pigments with similar properties have also been identified in *A. parasiticus* (Moreno *et al.*, 1987) and *P. cyclopium* (Ha-Huy-Ke & Luckner, 1979). The conidia of *P. patulum* exhibit a greenish grey pigmentation which may be of similar origin. Phylogenetic analysis of these sequences (figure 4.13) groups the minor *A. parasiticus* product with *A. nidulans wA* and the major product with *Colletotrichum lagenarium PKS1,* a gene involved in melanin biosynthesis. The pentaketide DHN melanins are extremely widespread among the ascomycetes and deuteromycetes and melanins of this type are certainly produced by *V. dahliae* (Wheeler, 1983). The sequenced PCR product from *V. dahliae* is extremely similar to the *PKS1* gene involved in DHN melanin biosynthesis in *Colletotrichum lagenarium* (Takano *et al.*, 1995), with an identity of 80.3 % at the nucleotide level and 92.1% at the amino acid level, and has probably been amplified from a homologue of this gene. Experiments using inhibitors of DHN melanin biosynthesis have indicated that green conidial pigments derived from pentaketide metabolites are widespread in the genera *Penicillium* and *Aspergillus* (Wheeler & Klich, 1995). Interestingly the same study found no evidence for these compounds in *A. parasiticus,* which is believed to synthesise its conidial pigment from the heptaketide intermediate parasperone A (Brown *et al.*, 1993).

Species	Polyketide	Class	Activity	Reference
1. *Penicillium patulum* IMI 92273 (ATCC 48925)	patulin	tetraketide	mycotoxin	Turner & Aldridge, 1983
	griseofulvin	heptaketide	antibiotic / antifungal	Turner & Aldridge, 1983
2. *Cytospora spp.* ATCC 20502	colletodiol	tri + tetraketide	antibiotic	O'Neill *et al.*, 1993
	grahamimycin	tri + tetraketide	antibiotic	Gurusiddaiah & Ronald, 1981
3. *Penicillium citrinum* NRRL 8084	compactin	nonaketide	inhibitor of cholesterol biosynthesis	O'Hagan 1991
	citrinin	pentaketide	mycotoxin, phytotoxin	Turner & Aldridge, 1983.
4. *Aspergillus parasiticus* ATCC 24690	aflatoxin(s)	decaketide	mycotoxin(s)	Turner & Aldridge, 1983
	parasperone A	heptaketide	spore pigment precursor	Brown *et al.*, 1993
5. *Phoma sp.* C2932	squalestatins	?	inhibitor of cholesterol biosynthesis	Dawson *et al.*, 1992
6. *Phoma etheridgei* UAMH 7003	phomalone	hexaketide	antifungal	Ayer & Jimenez, 1994
7. *Drechslera monoceras* ATCC 24641	monocerin	heptaketide	antifungal	Turner, 1971
	ravenelin	octaketide	pigment	Turner & Aldridge, 1983

Table 4.1 Polyketides produced by the fungal strains used in this study.

The latter finding may indicate that the major *A. parasiticus* LC1/2c PCR product, which is obviously closely related to *C. lagenarium PKS1,* has been amplified from an inactive pseudogene.

The LC3/5c oligonucleotides may have primed on a range of PKS genes involved in the biosynthesis of reduced polyketides. Alternatively, it is possible that these primers amplify homologues of *MSAS* from each fungal genome. The hybridisations to *P. patulum* genomic DNA seem to indicate a mixed PCR product from these primers, deriving from *MSAS* and the PKS gene designated *PKS2*. However, no evidence of expression of *PKS2* has been obtained (chapter 5) and it may be a pseudogene copy of *MSAS.* While not all of the species examined in this study are known patulin producers, this does not disqualify them from carrying the genes required for biosynthesis of 6-MSAS, although whether these genes would be properly expressed is another matter. Using a multiplex PCR method Geisen (1996) has found sequences homologous to aflatoxin biosynthesis genes in non-producing species of *Aspergillus* and *Penicillium.* The *MSAS* gene may also be involved in other biosynthetic pathways requiring 6-MSAS as an early intermediate.

Chapter 5

Sequence analysis of a putative PKS gene from *Penicillium patulum*

5.1 Introduction

Early attempts in this laboratory to screen genomic libraries for potential PKS genes using heterologous probes were largely unsuccessful and certainly inefficient (chapter 4 and Walsh, 1993). For this reason, degenerate PCR primers were designed with the aim of amplifying fragments from novel PKS genes for use as homologous probes. Primer sequences were derived from regions of *Penicillium patulum MSAS* and *Aspergillus parasiticus PKS1* encoding continuously conserved amino acid sequence. The two primers PKSUNI5A (GARCARATGGAYCCNCARCA) and PKSUNI3A (CCNGTRGCNCGNGTYACC) amplified a PCR product of 1.4 kb from *P. patulum* genomic DNA. When used to probe Southern blots of *P. patulum* genomic DNA, this PCR product hybridised strongly to bands corresponding to *MSAS* restriction fragments and also to another set of bands. Restriction analysis of this PCR product suggested that it was mixed, deriving from two different target sequences. The mixed product was blunt-end ligated into the vector pUBS1, and transformed into *E. coli.* Among the transformants three clones, designated pDK4/13/16, were identified as containing the fragment derived from a putative novel PKS gene. This cloned DNA was subsequently used as a homologous probe to screen a *P. patulum* genomic library, resulting in the isolation of several positive phage clones. The *MSAS*-derived probes CON1 and RED3 (figure 3.2) were then used to identify a 6.5 kb *Sst*I fragment from one of these phage clones that was likely to contain the complete *PKS2* sequence and which was subcloned for further analysis. This chapter reports the sequencing of the complete 6.5 kb fragment and analysis of the predicted gene product encoded therein.

5.2 Results

5.2.1 Sequence analysis of the 6.5 kb genomic fragment containing *PKS2*

The cloned chromosomal fragment was sequenced on both strands using a mixture of manual and automated methods (figure 5.1). Problems of sequence ambiguity were occasionally encountered in manual sequencing and were resolved either on the complementary strand, by designing new primers or by re-sequencing the uncertain region using automated methods. The cloned fragment of 6654 bp contains a single long open reading frame encoding a polypeptide of 1783 amino acids, similar in size to other fungal PKS enzymes such as MSAS (1774 residues). This predicted gene product is highly homologous to various PKS and FAS enzymes, in particular to MSAS (figure 5.2), and was designated PKS2.

The region of chromosome 3' of the stop codon predicted for *PKS2* was translated into all six possible reading frames and compared to sequences in the SwissProt database using the program BlastX (Gish & States, 1993) to check for the presence of genes linked to PKS2. Two shorter open reading frames on the opposite strand to *PKS2* were revealed by this similarity search, both encoding polypeptides highly homologous to regions of cytochrome p450 monooxygenase enzymes (figure 5.3).

5.2.2 Northern analysis of *PKS2* transcription

Total RNA from *P. patulum* mycelium and conidiating cultures of various ages was subjected to northern blot analysis to compare transcription of *PKS2* with transcription of *MSAS* (figures 5.4 and 5.5). Initial experiments showed that in total RNA from cultures that were expressing the *MSAS* transcript no mRNA of a similar size hybridised to the *PKS2* ketosynthase domain probe (figure 5.4). Subsequently a transcript of approximately 2.4 kb that hybridised weakly to the *PKS2* probe was observed in total RNA from mycelium and conidiating cultures (figure 5.5).

Figure 5.1 DNA sequence of a *P. patulum* genomic fragment containing a putative polyketide synthase gene *PKS2*.

The ketosynthase acyl-binding cysteine (**KS**), acyltransferase pantetheine-binding serine (**AT**), dehydratase active histidine (**DH**), ketoreductase NADPH binding site motif (**KR 1**) ketoreductase active site motif (**KR 2**) and acyl carrier protein phosphopantetheine-binding serine (**ACP**) are highlighted in bold type. Primer binding sites for PKSUNI5A (549-568 bp) and PKSUNI3A (1935-1952 bp) are underlined.

```
   1  GAGCTCCTATTGGAGTTAGAGGAAGTATAGACATAGCGAGGTCGTAGAATCTTCATATTA   60

  61  ATACATAGAGAGGAAGATTTATTGGTTTGAATTGACCACGAGGGAGATTAAAATCTTCAC  120

 121  ATCACCTTACTTAATCCATTGTCATAGAGTCTACATTACACAGCATAATGATTGCTTCAA  180
   1                                                MetIleAlaSerT    5

 181  CCAGCAGCACAGAGGCCCTTACCCCCACCGGGGATTCCAATAGTAAAGGCACAACGACTC  240
   6  hrSerSerThrGluAlaLeuThrProThrGlyAspSerAsnSerLysGlyThrThrThrP   25

 241  CCGCAACCAGCACCGAAGACACTGAGATGCAAGGTGATCTCAACCACAATGATACCCATG  300
  26  roAlaThrSerThrGluAspThrGluMetGlnGlyAspLeuAsnHisAsnAspThrHisA   45

 301  ATGACATTGCCATAATAGGGATGGCATGTCGCGTCCCAGGAGATGTGAAGTCCCCGTCTG  360
  46  spAspIleAlaIleIleGlyMetAlaCysArgValProGlyAspValLysSerProSerA   65

 361  CACTGTGGCAGTTCCTCCTCCAGAAAGGTGATGCTTCAGGCGACATGCCCTCCTGGCGTT  420
  66  laLeuTrpGlnPheLeuLeuGlnLysGlyAspAlaSerGlyAspMetProSerTrpArgT   85

 421  GGGATCCCTATCGACAGCGCCATCCTCGCAATGCGGCCGCACTGGCAAAAACAACGGCCA  480
  86  rpAspProTyrArgGlnArgHisProArgAsnAlaAlaAlaLeuAlaLysThrThrAlaL  105

 481  AGGGTTACTTCGTCAACGATATCGATCACTTCGATGCCGCCTTCTTTGCCATCTCCCCTC  540
 106  ysGlyTyrPheValAsnAspIleAspHisPheAspAlaAlaPhePheAlaIleSerProA  125

 541  GCGAGGCGGAGCAGATGGATCCCCAGCAGCGCATTGCGCTCGAGGTTGCCTGGGAGGCGT  600
 126  rgGluAlaGluGlnMetAspProGlnGlnArgIleAlaLeuGluValAlaTrpGluAlaL  145

 601  TGGAGAATGCAGGCATCTCTCCTTCCCGTCTTGCGGGCTCAGACACTTCAGTATACATGG  660
 146  euGluAsnAlaGlyIleSerProSerArgLeuAlaGlySerAspThrSerValTyrMetG  165

 661  GTGTCAACTCCGATGACTATGGCAAGCTGGTCTTGGAAGACCTCACCGGTGTCGGTGCAC  720
 166  lyValAsnSerAspAspTyrGlyLysLeuValLeuGluAspLeuThrGlyValGlyAlaH  185

 721  ATATGGGCGTTGGCACGGCCTACTGCGGAATCCCTAGCCGCATATCATATCTGCTAGATT  780
 186  isMetGlyValGlyThrAlaTyrCysGlyIleProSerArgIleSerTyrLeuLeuAspL  205

 781  TGATGGGTCCTAGCGTTGCACTGGATGCAGCCTGTGCATCATCCCTGGTGGCGGTACACC  840
 206  euMetGlyProSerValAlaLeuAspAlaAlaCysAlaSerSerLeuValAlaValHisH  225   → KS

 841  ATGCACGACAGGCTATCCGTGCTGGTGAGACCGATCTTGCCATCGCAGGTGGTGTCAACG  900
 226  isAlaArgGlnAlaIleArgAlaGlyGluThrAspLeuAlaIleAlaGlyGlyValAsnA  245

 901  CACTCTTGGGACCCGGGTTGACCAGGGTGCTGGATGAAGCGGGAGCTATCTCCGCGGATG  960
 246  laLeuLeuGlyProGlyLeuThrArgValLeuAspGluAlaGlyAlaIleSerAlaAspG  265

 961  GAAAGTGTCGTTCTTTCGATGACTCTGCTAACGGGTATGGCCGAGGCGAAGGAGCCGGTG 1020
 266  lyLysCysArgSerPheAspAspSerAlaAsnGlyTyrGlyArgGlyGluGlyAlaGlyV  285

1021  TGGTTATCCTAAAGCGTCTGGAAAAAGCTCTAACAGACGGTGACCGAGTTCTAGCTGTGC 1080
 286  alValIleLeuLysArgLeuGluLysAlaLeuThrAspGlyAspArgValLeuAlaValL  305

1081  TCAAGGGCAGTGCAGTGGCCTCAGATGGCAAGACGCTAGGAATTATGGCCCCCAATGCGC 1140
 306  euLysGlySerAlaValAlaSerAspGlyLysThrLeuGlyIleMetAlaProAsnAlaG  325

1141  AGGCACAGATCCTCGTTGCCCAAAAGGCACTCAAAGAGGCGAGGGTAACACCTGACTCCA 1200
 326  lnAlaGlnIleLeuValAlaGlnLysAlaLeuLysGluAlaArgValThrProAspSerI  345

1201  TCAGCTATATTGAGGCACATGCCACTTCCACCTCATTGGGAGATCCTACAGAGACGAGTG 1260
 346  leSerTyrIleGluAlaHisAlaThrSerThrSerLeuGlyAspProThrGluThrSerA  365

1261  CCCTAGCTGGAGTATATGGTGCTGGATCTGGACGGCATCCATGCAATCCGTGCTATATCG 1320
 366  laLeuAlaGlyValTyrGlyAlaGlySerGlyArgHisProCysAsnProCysTyrIleG  385

1321  GTTCCATCAAGCCCAATATCGGTCATTTAGAGGCTGGTGCCGGTGTCATGGGTCTCATCA 1380
 386  lySerIleLysProAsnIleGlyHisLeuGluAlaGlyAlaGlyValMetGlyLeuIleL  405

1381  AGGCGGTTCTTGTGCTTCGCCATGGCCAGGTCCCCCCGCAGGCCAATCTGCAGACGCTGA 1440
 406  ysAlaValLeuValLeuArgHisGlyGlnValProProGlnAlaAsnLeuGlnThrLeuA  425

1441  ATAGCAAAATCGCGTGGAAGAAGAGCTTACTATGCCCGGCCCGCGAGCTGGTGACATTGC 1500
```

```
 426  snSerLysIleAlaTrpLysLysSerLeuLeuCysProAlaArgGluLeuValThrLeuP   445

1501  CTCATGGTACCCCATCGCGACCATTGCGGGCCGCTGTAGCATCATACGGATATAGTGGCA  1560
 446  roHisGlyThrProSerArgProLeuArgAlaAlaValAlaSerTyrGlyTyrSerGlyT   465

1561  CAGTATCTCATGCAATTATCGAGGCCTTCGCAGGAGGATTCTTGTGCACCGATCGACCGG  1620
 466  hrValSerHisAlaIleIleGluAlaPheAlaGlyGlyPheLeuCysThrAspArgProA   485

1621  CCCAGCCTCCAGACAGTGATGGTGCACCAGTCCTCCTTCTCTTGTCCCTGCCTCAAGCGA  1680
 486  laGlnProProAspSerAspGlyAlaProValLeuLeuLeuLeuSerLeuProGlnAlaA   505

1681  ACCGAATATCAACAACAGCTGATGCCCTAAGCCAATGGCTGCGACATACTGATGGGGCGA  1740
 506  snArgIleSerThrThrAlaAspAlaLeuSerGlnTrpLeuArgHisThrAspGlyAlaI   525

1741  TATCCCTCGCAACGGTCGCCTCGACACTGTCACAGCGCCGCGCGCATCATCGCTTCCGCC  1800
 526  leSerLeuAlaThrValAlaSerThrLeuSerGlnArgArgAlaHisHisArgPheArgH   545

1801  ACGCCATTGTTGCTGATTCCGTCGCAAACGCCATCGCTACTCTCGATGATCTGTCCAAAA  1860
 546  isAlaIleValAlaAspSerValAlaAsnAlaIleAlaThrLeuAspAspLeuSerLysA   565

1861  ATGTGCCAAATCGCTGGGCAATCAACAACCGAATTGGAACCGAGGCAGCAAAAGGCCCGG  1920
 566  snValProAsnArgTrpAlaIleAsnAsnArgIleGlyThrGluAlaAlaLysGlyProV   585

1921  TGTGGGTATTCTCAGGCCATGGTGCGCAGTGGCCCGACATGGGTCGCGAATTATTTCACT  1980
 586  alTrpValPheSerGlyHisGlyAlaGlnTrpProAspMetGlyArgGluLeuPheHisS   605

1981  CTAGCCCGGTGTTCGGGGAAGTGGTGCGCAACCTGGAGCCAATTATTCAAGCAGAGCTAG  2040
 606  erSerProValPheGlyGluValValArgAsnLeuGluProIleIleGlnAlaGluLeuG   625

2041  GTTTCTCTGCAATTGAATCCCTGCAAGCTGGCTGTCTAAATCGCACTGATGTTGTTCAGG  2100
 626  lyPheSerAlaIleGluSerLeuGlnAlaGlyCysLeuAsnArgThrAspValValGlnA   645

2101  CAATGACATTTCTCATGCATCTGGGCATCGCCGCCGTGCTCGAGGCGGAGTCTGGTCCTC  2160
 646  laMetThrPheLeuMetHisLeuGlyIleAlaAlaValLeuGluAlaGluSerGlyProP   665

2161  CTACTGCTGTCGTGGGGCATTCCCTGGGCGAGGCTGCAGCGGCCGTCGTTTCTGGGGCTC  2220
 666  roThrAlaValValGlyHisSerLeuGlyGluAlaAlaAlaAlaValValSerGlyAlaL   685  → AT

2221  TAACTTGGCATGAGGGAGCCCTAGTAGTCTGTCGACGAGCGCGTCTTTACCGTGAGTTCA  2280
 686  euThrTrpHisGluGlyAlaLeuValValCysArgArgAlaArgLeuTyrArgGluPheI   705

2281  TCGGTGAAGGTGCGATGGCGCTGGTTCGGCTGCCAGCGTCTGAGGCCCGTGCGCGTATTG  2340
 706  leGlyGluGlyAlaMetAlaLeuValArgLeuProAlaSerGluAlaArgAlaArgIleA   725

2341  CCACGCACCTCGGCGCGTCGGTCGCAATCGAGGCATCACCAACCGTGTGCGTGATATCCG  2400
 726  laThrHisLeuGlyAlaSerValAlaIleGluAlaSerProThrValCysValIleSerG   745

2401  GTACTATTGACGCCGTTCAGAAGATCTCTCAAACATGGCGTGAAGAAGGTATCGAGGTGC  2460
 746  lyThrIleAspAlaValGlnLysIleSerGlnThrTrpArgGluGluGlyIleGluValA   765

2461  GTCCCGTCGCCACGGATGTGCCCTTCCATACACCACTCCTGGAGAAGCTAGCAGGTCCCC  2520
 766  rgProValAlaThrAspValProPheHisThrProLeuLeuGluLysLeuAlaGlyProL   785

2521  TCCGAGACGCGCTGAAAGGCGAGCTTCACCCTCAAGTACCGCGTCGGACGCTCTATTCGA  2580
 786  euArgAspAlaLeuLysGlyGluLeuHisProGlnValProArgArgThrLeuTyrSerT   805

2581  CTTCGCTGTTGGATCCACGTTTAGAAGTACCACGCGATGTCGAATACTGGGTGACGAATA  2640
 806  hrSerLeuLeuAspProArgLeuGluValProArgAspValGluTyrTrpValThrAsnM   825

2641  TGATCCAACCGGTCCGACTGCAGTCGACAGTGGCCGCACTCGTGGACGACGGATTCCGCG  2700
 826  etIleGlnProValArgLeuGlnSerThrValAlaAlaLeuValAspAspGlyPheArgA   845

2701  CATTTGTGGAGCTGGCGAGTCATCCCATAATAACACACTCGATAGTCGAGACTGTCAGCG  2760
 846  laPheValGluLeuAlaSerHisProIleIleThrHisSerIleValGluThrValSerG   865

2761  AGCGGACAACCGATCGGTTTATGGTGACACCCACAATGTTGCGTAAGCAGCCGGTACTCA  2820
 866  luArgThrThrAspArgPheMetValThrProThrMetLeuArgLysGlnProValLeuL   885

2821  AGAGCATCTTGGCTGCCGTGGGCCGTCTGCACTGCTTCGGCTGCACTGTGAAATACACAG  2880
 886  ysSerIleLeuAlaAlaValGlyArgLeuHisCysPheGlyCysThrValLysTyrThrA   905

2881  ATCTCGACCCGGCTGCGcCATGGAGTTCATCTGTGCCCGGTACAATTTGGAATCGCCAGC  2940
 806  hrSerLeuLeuAspProArgLeuGluValProArgAspValGluTyrTrpValThrAsnM   825

2641  TGATCCAACCGGTCCGACTGCAGTCGACAGTGGCCGCACTCGTGGACGACGGATTCCGCG  2700
 826  etIleGlnProValArgLeuGlnSerThrValAlaAlaLeuValAspAspGlyPheArgA   845

2701  CATTTGTGGAGCTGGCGAGTCATCCCATAATAACACACTCGATAGTCGAGACTGTCAGCG  2760
 846  laPheValGluLeuAlaSerHisProIleIleThrHisSerIleValGluThrValSerG   865

2761  AGCGGACAACCGATCGGTTTATGGTGACACCCACAATGTTGCGTAAGCAGCCGGTACTCA  2820
 866  luArgThrThrAspArgPheMetValThrProThrMetLeuArgLysGlnProValLeuL   885

2821  AGAGCATCTTGGCTGCCGTGGGCCGTCTGCACTGCTTCGGCTGCACTGTGAAATACACAG  2880
 886  ysSerIleLeuAlaAlaValGlyArgLeuHisCysPheGlyCysThrValLysTyrThrA   905

2881  ATCTCGACCCGGCTGCGcCATGGAGTTCATCTGTGCCCGGTACAATTTGGAATCGCCAGC  2940
```

```
 906  spLeuAspProAlaAlaProTrpSerSerSerValProGlyThrIleTrpAsnArgGlnP   925

2941  CGTTCTACCGCGCTGTGAGCACAATGACGGCTGCACAGTTGGCGCCAACACACAAACCTG  3000
 926  roPheTyrArgAlaValSerThrMetThrAlaAlaGlnLeuAlaProThrHisLysProA   945

3001  CCGCAAATGACCTGCTAGGTACTCGTACCGCGCTGTGGGGAACGGACGAGGTGCTTTATC  3060
 946  laAlaAsnAspLeuLeuGlyThrArgThrAlaLeuTrpGlyThrAspGluValLeuTyrG   965

3061  AAACCCGCCTCGAGGAAGACAATCGACCTTTTCCCGGCCGTCACCCACTGCATGGATCAG  3120
 966  lnThrArgLeuGluGluAspAsnArgProPheProGlyArgHisProLeuHisGlySerG   985  → DH

3121  AGATAGTACCTGCCGCTGTGCTATTGCATACATTTTTGCGCGCACTCTCTCCTAGTAGCG  3180
 986  luIleValProAlaAlaValLeuLeuHisThrPheLeuArgAlaLeuSerProSerSerV  1005

3181  TGGAGGATGTGTCGCTGCTGGTGCCTGTAGTGGTGTCTCCGGCCCGTGAGATCCAGATCC  3240
1006  alGluAspValSerLeuLeuValProValValValSerProAlaArgGluIleGlnIleA  1025

3241  GGCATAACACGCGTGAGATCGCTATCACATCTCGCCTGGAGGAGTCTACGAGCAACGAAG  3300
1026  rgHisAsnThrArgGluIleAlaIleThrSerArgLeuGluGluSerThrSerAsnGluA  1045

3301  ACGGCTCGTGGCTTGTCAATACCATCGCGACTAGGGGCGCTACTGATGTAGCAACCTCCA  3360
1046  spGlySerTrpLeuValAsnThrIleAlaThrArgGlyAlaThrAspValAlaThrSerI  1065

3361  TCAGCCATATCAACATAGCTGAAATCAGGAAAAGACTGCCACAGAAGCTGTCGGACAACT  3420
1066  leSerHisIleAsnIleAlaGluIleArgLysArgLeuProGlnLysLeuSerAspAsnP  1085

3421  TCTCAATGGATTACCTCGCCTCTGTAGGCGTGTCTGCCATGGGATTTCCATGGCGGGTAA  3480
1086  heSerMetAspTyrLeuAlaSerValGlyValSerAlaMetGlyPheProTrpArgValT  1105

3481  CTCAGCATGTGGCCAGTGACGACGAGATGCTGGCCATGGTAAACGCCAATCCGGACAACC  3540
1106  hrGlnHisValAlaSerAspAspGluMetLeuAlaMetValAsnAlaAsnProAspAsnL  1125

3541  TACCTGGCATGGACGATTTTTTGGTGTCCGTCATGGATGCATCAACGTCCATCGCATCAA  3600
1126  euProGlyMetAspAspPheLeuValSerValMetAspAlaSerThrSerIleAlaSerT  1145

3601  CCCTCTGGCATTGTATGCCCCGCCTCCGTATGCCCACAGCTGTCCGTCGCATCGTGGCCG  3660
1146  hrLeuTrpHisCysMetProArgLeuArgMetProThrAlaValArgArgIleValAlaV  1165

3661  TTGATGTAGCCACCCCACAAGTCGTGTACATCCACTGCAGCAAGGCAAAATCGTCCGTCG  3720
1166  alAspValAlaThrProGlnValValTyrIleHisCysSerLysAlaLysSerSerValA  1185

3721  ACGCAGCTGATGTGATAATCAGTAGTGAGGATGGTACCGTTCTGATGGAGATCCAGGGAA  3780
1186  spAlaAlaAspValIleIleSerSerGluAspGlyThrValLeuMetGluIleGlnGlyM  1205

3781  TGGCCTTCGCCGGTGTCGAGGGCGACTCGTTGTCACGAAAAAATACCTCTGGCCTCGTGC  3840
1206  etAlaPheAlaGlyValGluGlyAspSerLeuSerArgLysAsnThrSerGlyLeuValH  1225

3841  ACCAAATCAGCTGGCCACCGGCCGCTCTGGTTGAGGAGCCGTTAGAATTTTCCCACGTCG  3900
1226  isGlnIleSerTrpProProAlaAlaLeuValGluGluProLeuGluPheSerHisValA  1245

3901  CCTTTCTCGCACCAGATGCCGCGACGACTTACGTGGAAACCTACCAGATGCAGCTTGAGA  3960
1246  laPheLeuAlaProAspAlaAlaThrThrTyrValGluThrTyrGlnMetGlnLeuGluS  1265

3961  GCAGAGGTATCTCCACGTCTATCCATGATTATGCCTCAGACTTGCCCTTGACAGCTCATT  4020
1266  erArgGlyIleSerThrSerIleHisAspTyrAlaSerAspLeuProLeuThrAlaHisS  1285

4021  CGTCTATGGCGGTTGTCTATCTCCCTCAAGTCACCGATCAAATTTTCGAGACCGCCACCA  4080
1286  erSerMetAlaValValTyrLeuProGlnValThrAspGlnIlePheGluThrAlaThrS  1305

4081  GCGCCTGCAGTCGCCTCGTATCGGCAGCCCAGTTTATTTTATCCTTCTCGAAAAAACCCA  4140
1306  erAlaCysSerArgLeuValSerAlaAlaGlnPheIleLeuSerPheSerLysLysProT  1325

4141  CCGTCCGCCTCTTCACTCTCACCTCCGAAACCAATCTAGGCCACAGCGCCCTCACAGGCC  4200
1326  hrValArgLeuPheThrLeuThrSerGluThrAsnLeuGlyHisSerAlaLeuThrGlyL  1345

4201  TCGGCCGGATTCTGCACACCGAACACCCCGAGATCTGGGGTAGTGTCATCGATATTGAAG  4260
1346  euGlyArgIleLeuHisThrGluHisProGluIleTrpGlySerValIleAspIleGluA  1365

4261  ATCCCTCGGTCTTCCCGCTTATGGCAATGCGCTATGTGCGTAATACTGACATTATCAAGA  4320
1366  spProSerValPheProLeuMetAlaMetArgTyrValArgAsnThrAspIleIleLysI  1385

4321  TCCAGGATGGGGTCCCGCGCACTGCATGCCTCCGTCCGCTCCGCTCCACCCCGCTCCACC  4380
```

```
1386  leGlnAspGlyValProArgThrAlaCysLeuArgProLeuArgSerThrProLeuHisP   1405
                .         .         .         .         .         .
4381  CCGCTTCTGGACCCGCTGCTCTCACTTTCTCCTTGGCATCCACGTATCTAATCACCGGTG   4440
1406  roAlaSerGlyProAlaAlaLeuThrPheSerLeuAlaSerThrTyrLeuIleThrGlyG   1425
                .         .         .         .         .         .
4441  GACTGGGCTGTCTCGGTCTCTCCGTAGCCGAGTGGATGGTCACTCAAGGCGCGAAACGCA   4500
1426  lyLeuGlyCysLeuGlyLeuSerValAlaGluTrpMetValThrGlnGlyAlaLysArgI   1445 → KR 1
                .         .         .         .         .         .
4501  TTCTGCTACTCTCTCGTCGCTCACTACCCCCGAGGTCCACCTGGACCACCTCGCATAAAc   4560
1446  leLeuLeuLeuSerArgArgSerLeuProProArgSerThrTrpThrThrSerHisLysP   1465
                .         .         .         .         .         .
4561  CCGAGACCTCGTCTATCATTCAAAGTATCCTCTCTCTTGAACGCCTTGGTGCCACTATCC   4620
1466  roGluThrSerSerIleIleGlnSerIleLeuSerLeuGluArgLeuGlyAlaThrIleH   1485
                .         .         .         .         .         .
4621  ACACCCTAGCCATCGATATTTCCCATCCATCCGCTGTGGTGAGTCTTCGCTCGGCGCTCA   4680
1486  isThrLeuAlaIleAspIleSerHisProSerAlaValValSerLeuArgSerAlaLeuT   1505
                .         .         .         .         .         .
4681  CCACACTCAACCTCCCGCCCGTTGCTGGCGTAGTTCACGCCGCTGGCATAATCCACGACC   4740
1506  hrThrLeuAsnLeuProProValAlaGlyValValHisAlaAlaGlyIleIleHisAspG   1525
                .         .         .         .         .         .
4741  AGCTCATCGGGCAAATTACTCCAGATGTGTTTGATGCCGTCCTTGCACCGAAGATCGCAG   4800
1526  lnLeuIleGlyGlnIleThrProAspValPheAspAlaValLeuAlaProLysIleAlaG   1545
                .         .         .         .         .         .
4801  GCGCCTTGGTGTTGGACACCGTCTTCCCTCCCAATTCCCCTGAGCTGGACTTCTTCATCC   4860
1546  lyAlaLeuValLeuAspThrValPheProProAsnSerProGluLeuAspPhePheIleL   1565
                .         .         .         .         .         .
4861  TCTTCTCTTCTTGCGGTCAACTCTTGGGATTCCCGGGTCAGGCGTCATACGCGAGTGGAA   4920
1566  euPheSerSerCysGlyGlnLeuLeuGlyPheProGlyGlnAlaSerTyrAlaSerGlyA   1585 → KR 2
                .         .         .         .         .         .
4921  ATTCGTTTTTGGACGCCTTGGCGCGGTCACGCAGAAAGGAGGGTGATAATTCCATCTCTC   4980
1586  snSerPheLeuAspAlaLeuAlaArgSerArgArgLysGluGlyAspAsnSerIleSerL   1605
                .         .         .         .         .         .
4981  TTCTTTGGACGTCTTGGCGGGGCATGGGTATGGGAGCTAGTTCTAATGGAGCCCTAGAGG   5040
1606  euLeuTrpThrSerTrpArgGlyMetGlyMetGlyAlaSerSerAsnGlyAlaLeuGluA   1625
                .         .         .         .         .         .
5041  CGGAACTCTATGCTCGTGGAATTACAGATGTGACTCCTGAAGAGGCGTTCCGGGCTTGGA   5100
1626  laGluLeuTyrAlaArgGlyIleThrAspValThrProGluGluAlaPheArgAlaTrpS   1645
                .         .         .         .         .         .
5101  GTGCTATCTCTGCAGTTGAGGGTGCCGATCATGGAGTTGTGCTTCGTGCTCGGCCGCTTG   5160
1646  erAlaIleSerAlaValGluGlyAlaAspHisGlyValValLeuArgAlaArgProLeuG   1665
                .         .         .         .         .         .
5161  AATCTCGGGAACCGCTGCCACATGCGATTCTGAGAGATATTGTCACGAGGAAAGAGAGAA   5220
1666  luSerArgGluProLeuProHisAlaIleLeuArgAspIleValThrArgLysGluArgI   1685
                .         .         .         .         .         .
5221  TTGGGGATGACGAGAGTGGGGAAAAGGAAAAACAAAGGAACAAATTGACTGGCAAAGGAT   5280
1686  leGlyAspAspGluSerGlyGluLysGluLysGlnArgAsnLysLeuThrGlyLysGlyL   1705
                .         .         .         .         .         .
5281  TGGCGGAACATTTGCGTGTCGTTGTGAAAAAGTGTGTGTCGACGACATTATCGATCCCTG   5340
1706  euAlaGluHisLeuArgValValValLysLysCysValSerThrThrLeuSerIleProG   1725
                .         .         .         .         .         .
5341  AAGATGAAGTAGACGAGACAGTGGCGTTGCCAGAGATGGGCATGGACTCTGTCATGACAG   5400
1726  luAspGluValAspGluThrValAlaLeuProGluMetGlyMetAspSerValMetThrV   1745 → ACP
                .         .         .         .         .         .
5401  TGAACTTCCGCATGAGCCTGCAGCAGACGCTGATGGTGGGTGTTGGGCCAACGTTGGTAT   5460
1746  alAsnPheArgMetSerLeuGlnGlnThrLeuMetValGlyValGlyProThrLeuValT   1765
                .         .         .         .         .         .
5461  GGAAGTACCCTACGATACATCATCTGGTTGAATATTTCTGTCAGGTATTGGAAGAATAGT   5520
1766  rpLysTyrProThrIleHisHisLeuValGluTyrPheCysGlnValLeuGluGluEnd   1784
                .         .         .         .         .         .
5521  TTTAGCTATCGTTCCATTGGAGACTTGTATATAGTTCTATTTTACTTCTTCTTCTAATCT   5580
                .         .         .         .         .         .
5581  TGCAACTCGGAATATAAATCATATCACTAACTATAAAGACAGATGGTCTAATAAAGGCAT   5640
                .         .         .         .         .         .
5641  ATCAATCTATAATTTGTAGGACATAACATCGCCAACTATAGAATCAACCACGTAGAGACA   5700
                .         .         .         .         .         .
5701  TCTCCATCTATACCTAATCTTCTCTCCGTTTCGTCAATGCTAGGAACTCTCCGTGAAACT   5760
                .         .         .         .         .         .
5761  GGGGAACAACCATATCATCCCACTTCATATCCTCCACGGTCGTTCCTACTAGGTCCAAAT   5820
                .         .         .         .         .         .
5821  CGAATCGTCGCACAATATATGCAATGGCCATATACATCTCCATATACGCCATACTAAACA   5880
                .         .         .         .         .         .
5881  ACAAATTCGTTAGTCACATTTATTTCATGAACCAAGGGGGGCACATACTTGATCCCCAGA   5940
```

```
5941  CAAGACCGACTCCCTCGTGAAAACGGTACTAGAAATTTAGCTCTCTCTGCTGTCACCTCG  6000
6001  GACCCCTGGAGCCATCGCTCTGGATCAAACACACGAGGATTCTCGAAAAGCGTCTCATTG  6060
6061  TAGTGAATATGGTAAATGGCCATGCCTACCCGAGTTCCCGCAGGGATAAACCGGCCGTCT  6120
6121  GGAAGACGGACACCACCTGGCGGAACAAGACGGGGAGTGCGAGAGGCTGCGGGAGAGGAA  6180
6181  TATCTGAGGCCCTCTTTTATTACTGCAGTCTGTTTTGTGCCATATAAGCGTTAGTTCAAT  6240
6241  TGTTTTTAGGAGTAAGAATAGGATACAAAGTACGGACCAGATATGGTAGCGCTTCAAGCT  6300
6301  GAATATACGAGACATGAGAATTGGACGTCGGCATGACAGTTTTCAGTTCCGCTAGAAGTC  6360
6361  GGTCATGGATTTCACGGTCATTGAAGATATTGAAAAGAGTGACTGTGTTGGCAATTCCGG  6420
6421  TGGTATCGTCCCCAGCGGAGATCACGGTAAGAAGCTCATCACGGAGACCCTGTTTGTCCG  6480
6481  GTACTGCATAACCCTTCTCAGGAAGCGGAATGAGCAGCCTCTCCATGAGAGTCGGGAAGG  6540
6541  CAGGGGTTTTGCCCGCTGCTGCGGCGGCGATATATTCATCAATGCGACGAGCAGCGAACT  6600
6601  GGTAGAACTAATTAGTATAGATGAGGTTGCTGACGGAACGTATAAGTCGAGCTC  6654
```

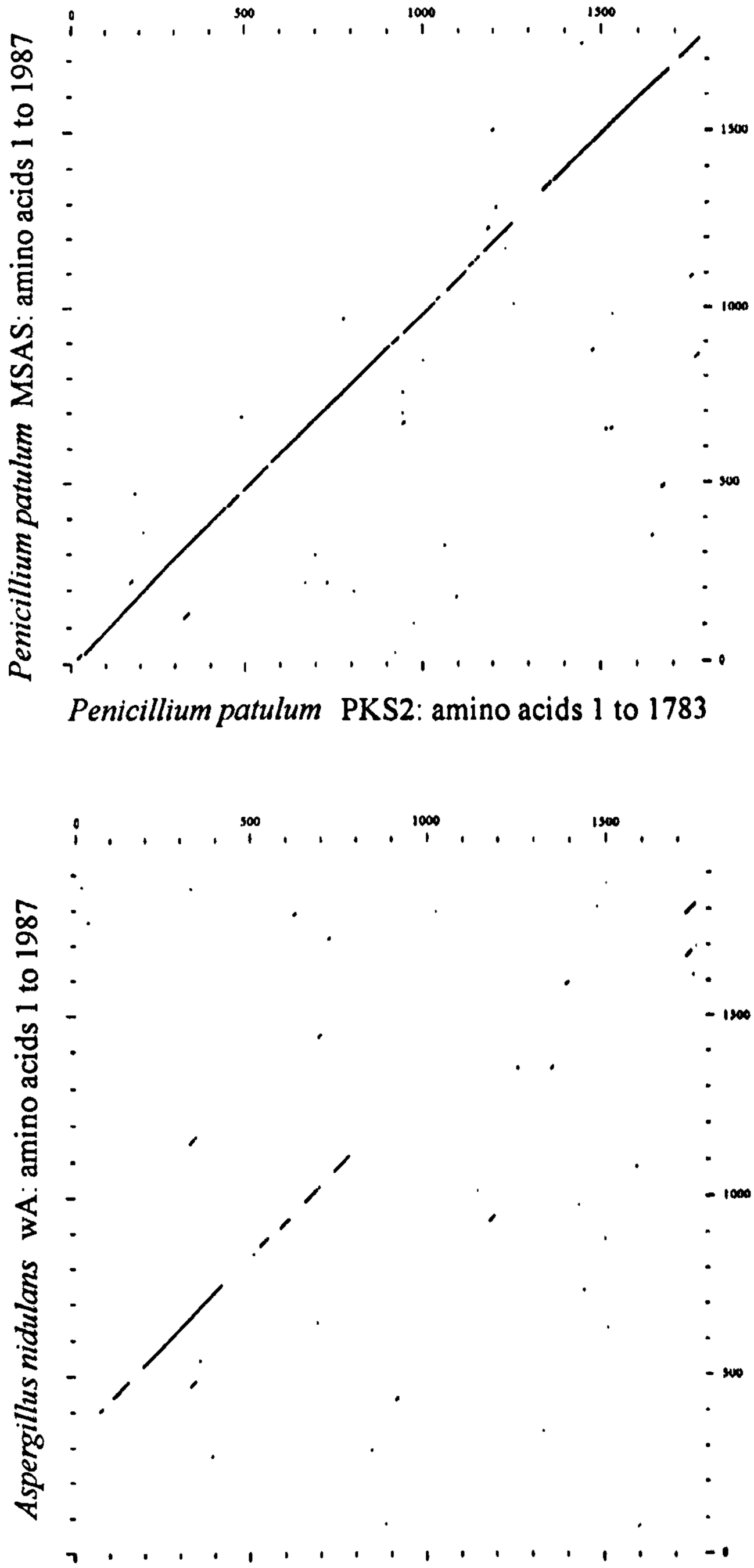


Figure 5.2 Matrix comparison of predicted gene product *P. patulum* PKS2 with *A. nidulans* wA. The matrix data files were calculated using a window size of 30 and a stringency of 25 with the program COMPARE and plotted using the program DOTPLOT (Devereux *et al.*, 1984).

Figure 5.3 BlastX search of the SwissProt database indicates the presence of a Cytochrome P450 gene downstream of *P. patulum PKS2*.

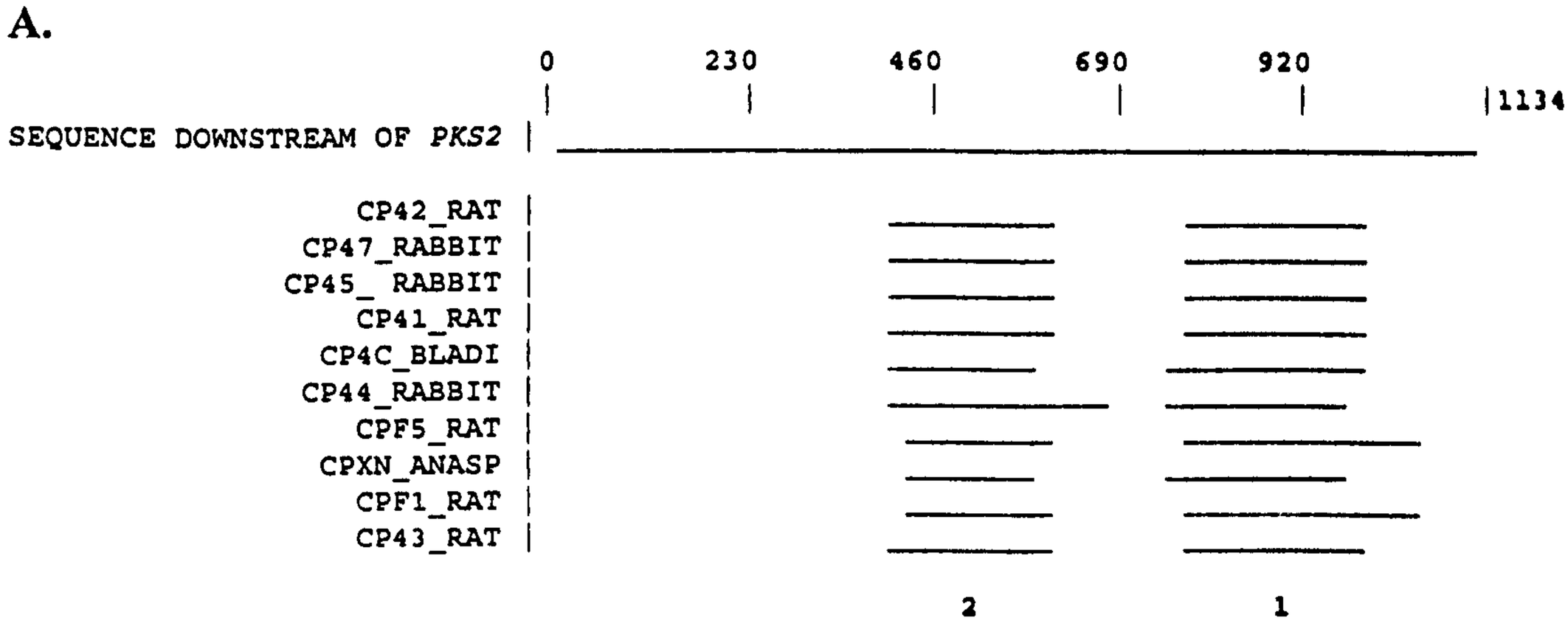
A. Query sequence and locally aligned high-scoring regions from database sequences

The upper line represents the 1134 bp of chromosome immediately downstream of *PKS2*. This query sequence has been translated into all six reading frames and used to search the SwissProt database using the BlastX program (Gish & States, 1993); beneath this are shown the positions of the two aligned regions in the ten highest-scoring protein sequences (HSPs). All are cytochrome p450 sequences (identity codes shown to the left) and are aligned to translations of reading frames on the complementary strand to *PKS2*.

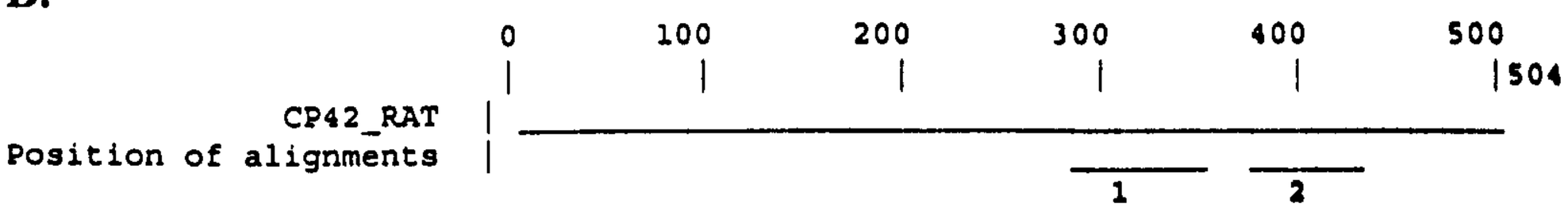
B. High scoring local alignments of CP42_RAT cytochrome p450 IVA2 (lauric acid ω-hydroxylase) to translations of complementary strand of query sequence.

The upper diagram shows the position of the alignments **1** and **2** on the complete CP42_RAT polypeptide (504 amino acids). A consensus sequence is shown between the query sequence (PKS2 downstream) and the cytochrome p450 sequence; amino acid similarities are indicated by + . The conserved heme binding motif **FXXGXXXCXG** is highlighted in bold type.

A.



B.



1. Score = 69, Number expected by chance = 2.2e-10
   Identities = 21/82 (25%), Similarities = 41/82 (50%), Frame = -3

PKS2 DOWNSTR: 1006 MERLLIPLPEKGYAVPDKQGLRDELLTVISAGDDTTGIANTVTLFNIFNDREIHDRLLAE 827
++ LL E G ++ D+ LR E+ T + G DTT + + + E +R E
CP42_RAT: 286 LDILLFAKMEDGKSLSDED-LRAEVDTFMFEGHDTTASGISWVFYALATHPEHQERCREE 344

PKS2 DOWNSTR: 826 LKTVMPTSNSHVSYIQLEALPY 761
+++++ S V++ L+ +PY
CP42_RAT: 345 VQSILGDGTS-VTWDHLDQMPY 365

2. Score = 148, Number expected by chance = 2.2e-10
   Identities = 29/69 (42%), Similarities = 44/69 (63%), Frame = -1

PKS2 DOWNSTR: 612 VRLPDGRFIPAGTRVGMAIYHIHYNETLFENPRVFDPERWLQGSEVTAERAKFLVPFSRG 433
V PDGR IP G RV + IY +H+N + + NP+VFDP R+ S + + +PFS G
CP42_RAT: 391 VTFPDGRSIPKGIRVTILIYGLHHNPSYWPNPKVFDPSRF---SPDSPRHSHAYLPFSGG 447

PKS2 DOWNSTR: 432 SRSCLGIKY 406
+R+C+G ++
CP42_RAT: 448 ARNCIGKQF 456

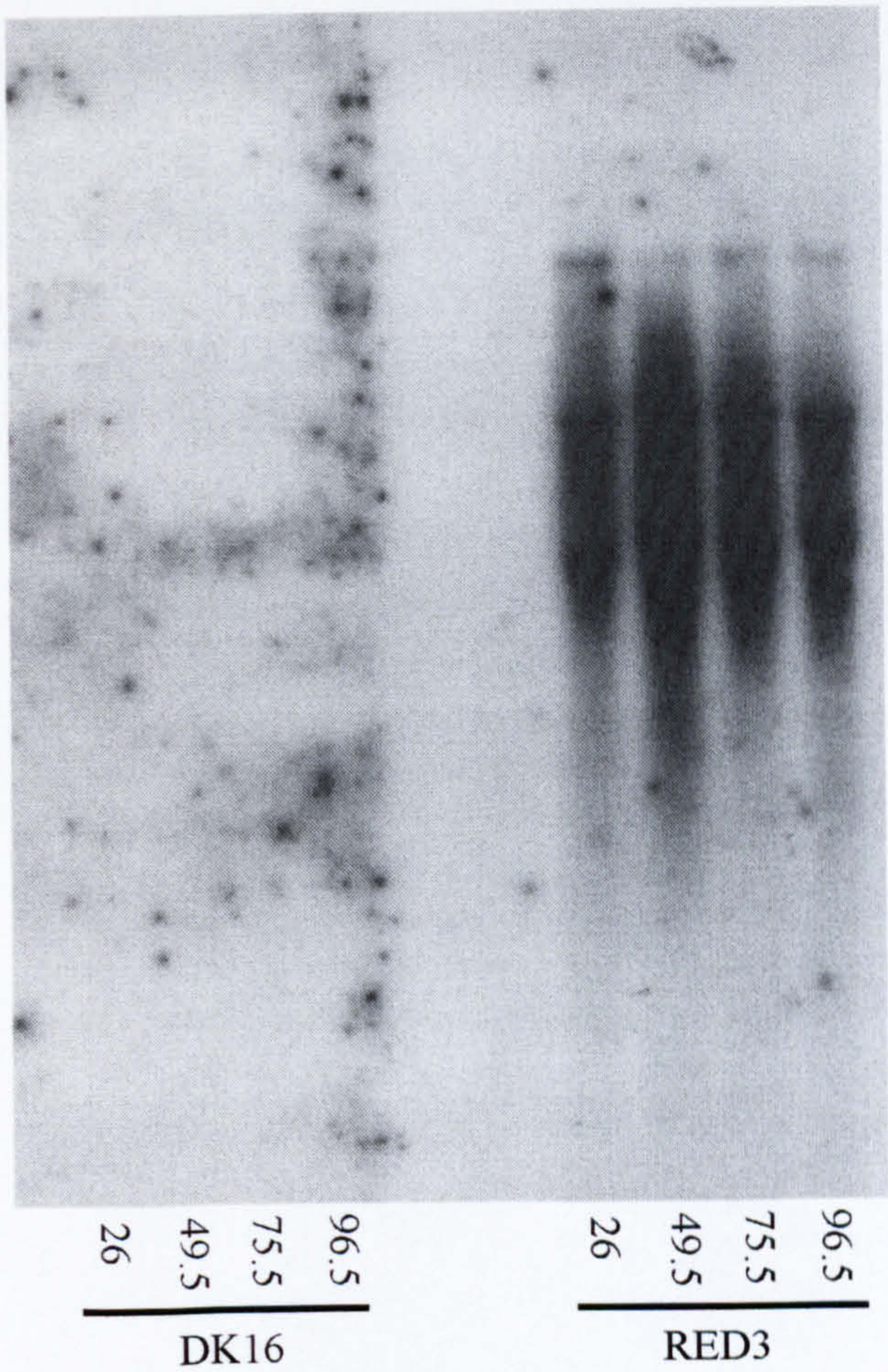


Figure 5.4 Northern blot of *Penicillium patulum* total RNA showing that *PKS2* is not transcribed under the same culture conditions as *MSAS*.

Total RNA was extracted from mycelium of different ages (indicated as number of hours post-inoculation beneath each lane), fractionated on duplicate formaldehyde denaturing agarose gels and blotted onto nylon membranes. These northern blots were probed with ^{32}P-labelled DNA as follows: **RED3**, *MSAS* ketoreductase domain restriction fragment (see figure 3.2); **DK16**, *PKS2* ketosynthase domain PCR product (see section 5.1). Bands corresponding to the *MSAS* transcript and degraded *MSAS* transcript co-migrating with 28S and 18S ribosomal RNAs are indicated by arrows.

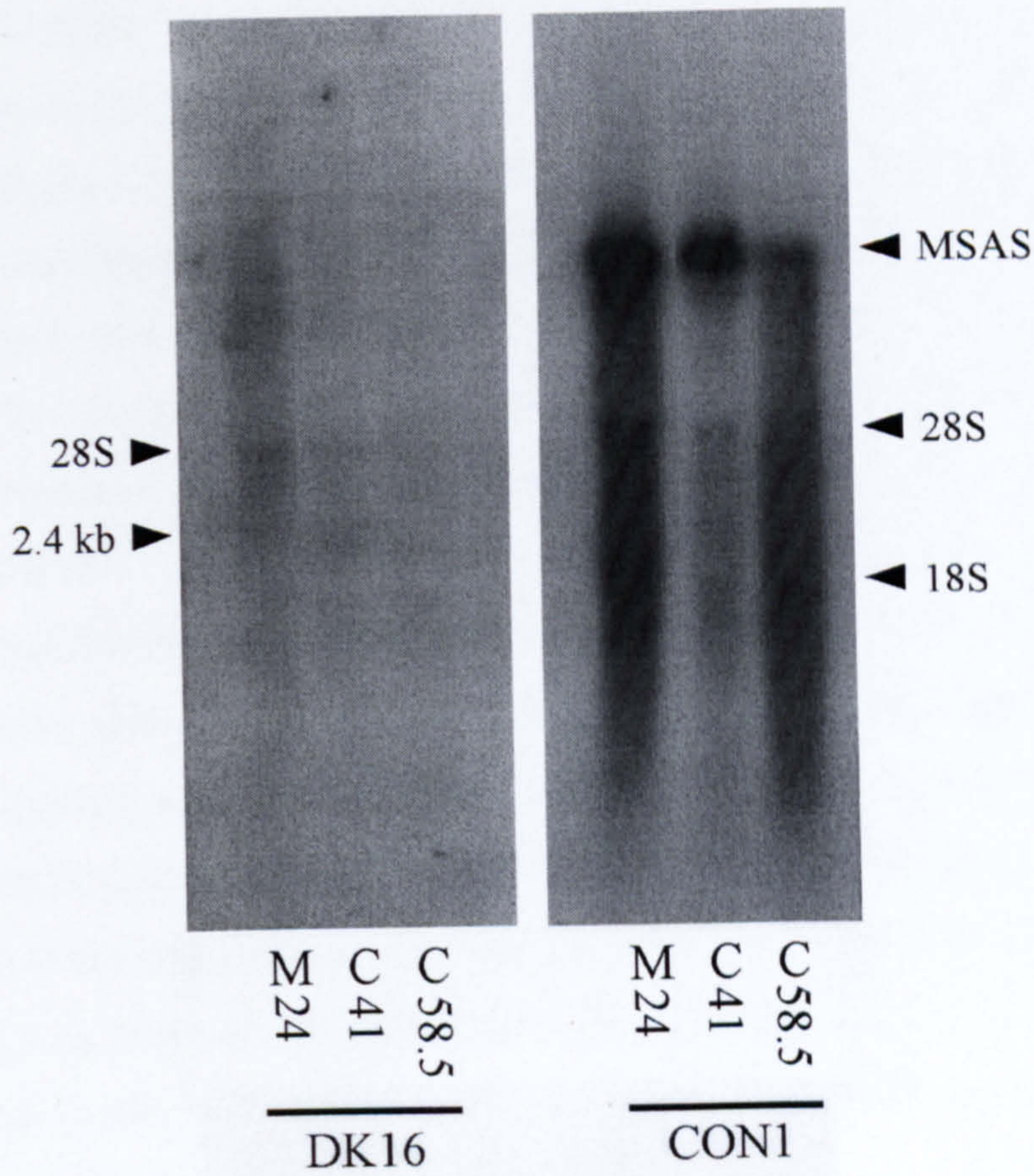


Figure 5.5 Northern blot of *Penicillium patulum* total RNA extracted from mycelium and conidiating cultures showing weak hybridisation of *PKS2* ketosynthase domain to 2.4 kb transcript.

Total RNA was extracted from 24 h mycelium, 41 h and 58.5 h conidiating cultures (indicated beneath each lane), fractionated on duplicate formaldehyde denaturing agarose gels and blotted onto nylon membranes. These northern blots were probed with ^{32}P-labelled DNA as follows: **CON1**, probed with *MSAS* ketosynthase domain probe (see figure 3.2); **DK16**, *PKS2* ketosynthase domain PCR product (see section 5.1). Bands corresponding to the *MSAS* transcript and to transcripts co-migrating with 28S and 18S ribosomal RNA are indicated by arrows.

5.3 Discussion

5.3.1 Transcriptional control motifs and predicted sites of initiation and termination of translation for *PKS2*

Only a limited region (167 bp) 5' of the putative start of translation for *PKS2* was cloned and sequenced (figure 5.1). This region has been examined for the presence of common transcriptional control motifs (see section 3.3.3 for a discussion of the functional significance of these motifs). Within this distance from the start of translation one might expect to find a TATA box (Unkles, 1992); the closest match in this region is TATAGA, situated 142 bp before the predicted start of translation (figure 5.1). No CT box can be identified 5' of *PKS2*.

The predicted sites for initiation and termination of translation for *PKS2* are based on the assumption that the gene consists of a single large open reading frame and are consistent with the homology between *PKS2* and other "MSAS-type" genes. The predicted start of translation for *PKS2* is in the context of the sequence AGCATAATGA, a very rough appproximation to the consensus for a eukaryotic initiation site GCCRCCAUGG (Lewin, 1997; section 3.3.5). However the predicted initiation sequence for *PKS2* does possess an A at the important position -3 with respect to the AUG codon, as found in 64% of filamentous fungi (Gurr *et al.*, 1987; Unkles, 1992).

The predicted stop codon for *PKS2* is followed by a pyrimidine (thymine), found in only 30% of filamentous fungal genes (Unkles, 1992). Downstream 110 bp from this stop codon is a perfect AATAAA polyadenylation signal motif (see section 3.3.4).

5.3.2 Analysis of amino acid sequence around the predicted active sites of PKS2

Five enzymic domains have been identified in the predicted PKS2 gene product by sequence homology to other PKS and FAS enzymes, more specifically by the presence of active site motifs known to be conserved in PKS and FAS enzymes (figure 5.6). Dot matrix comparison of the predicted gene products PKS2 and MSAS of *P. patulum* (figure 5.2) indicates almost continuous homology along the lengths of the proteins, suggesting that PKS2 possesses the same five enzymic domains as MSAS. As discussed in chapter 4, the known fungal PKS genes appear to fall into two subgroups: the "MSAS-type and the "wA-type". The dot matrix comparisons of PKS2 with representative gene products from each of

these subgroups (figure 5.2) clearly indicate that sequence homology between PKS2 and *A. nidulans* wA is much lower than that between PKS2 and MSAS and is limited mainly to the ketosynthase domain, which is highly conserved between all type I PKS and FAS (Aparicio *et al.*, 1996). This places PKS2 in the MSAS-type subgroup, which is consistent with the phylogenetic analysis of PKS ketosynthase domains discussed in chapter 4. Local frequency of non-conservative mutation, measured by the comparison of two homologous gene product sequences has been used as an indicator of domain boundary location in the animal FAS (Witkowski *et al.*, 1991). A modified approach was used to assess likely domain boundaries in PKS2 by comparison of this protein sequence with MSAS (figure 5.7). The rationale behind this analysis is that interdomain regions may primarily serve a linker or spacer function and may be less highly conserved than the functional domains (Donadio & Katz, 1992; Witkowski *et al.*, 1991). Therefore the regions of highest homology between these two amino acid sequences may represent the extent of the enzymic domains while areas of lower homology may represent linker or spacer regions. The predicted limits of each domain were correlated with the consensus sequences identified at the domain boundaries of the EryA and EryB gene products (Bevitt *et al.*, 1992; Donadio & Katz, 1992). Identification of the boundaries of each domain will facilitate comparison of PKS2 with other type I PKS enzymes and may allow the over-expression of single domains in heterologous systems, a prerequisite for analysis of domain structure and function. Knowledge of the extent of each functional domain will also be required for the design of recombinant fungal PKS enzymes which may be used in the biosynthesis of novel polyketide metabolites.

5.3.2.1 Ketosynthase domain

The ketosynthase active site consensus sequence **GPxxxxxxxCxSxL** (Cortes *et al.*, 1990) can be identified in the deduced PKS2 amino acid sequence around Cys_{216}. The amino acid sequence around the active cysteine is highly conserved with the two MSAS sequences from *P. patulum* and *A. terreus,* and PKS1 from *A. parasiticus*. It is also conserved to a lesser, but significant degree with other type I PKS and FAS ketosynthase enzymes (figures 5.5 and 4.12). Both of the histidine residues noted by Aparicio *et al.* (1996) as possibly involved in increasing the nucleophilicity of the active site cysteine are conserved in PKS2 at His_{351} and His_{393}. Conserved sequence motifs have been identified at the N-terminal and

C-terminal boundaries of the ketosynthase domain: **PiAiVgmaCR** and **GTNAHvIeE** respectively, where upper case letters represent invariant residues (Donadio & Katz, 1992). These motifs can be clearly identified in PKS2 at positions 46-55 and 464-472, suggesting a domain size of 427 amino acids which is similar to the 425 amino acids assigned to this domain in the EryA proteins (Donadio & Katz, 1992).

5.3.2.2 Acyltransferase domain

Polyketide and fatty acid synthase complexes require acyltransferase activity to catalyse the transfer of acyl-CoA ester extender units onto the acyl carrier protein. The presence of the signature motif **GHSxG** around the catalytic residue Ser_{672} identifies an acyltransferase domain within the PKS2 polypeptide. An arginine residue which is positionally conserved in all multifunctional PKS and FAS enzymes has been identified as playing an important role in the binding of malonyl and methylmalonyl moieties to the transacylase domain (Rangan & Smith, 1997; Serre *et al.*, 1995). This residue is present in PKS2 along with the other universally conserved active site residue, His_{774} (Serre *et al.*, 1995). Other conserved active site residues that can be identified in the acyltransferase domain of PKS2 are Gln_{594}, Gln_{644}, Phe_{773}, Thr_{805}, Asn_{824} and Val_{829} (Serre *et al.*, 1995), all of which are positionally conserved in the other "MSAS-type" enzymes.

In the fungal PKS, along with the FAS and actinomycete type II PKS enzymes, the extender unit is usually malonyl-CoA (see section 1.3.1). In contrast, the modular actinomycete PKS enzymes responsible for biosynthesis of reduced or complex polyketides utilise both malonyl and methylmalonyl-CoA extender units. The presence of a separate AT domain for each extension step in these modular enzymes has enabled the identification of divergent sequence motifs specific for both malonyl and methylmalonyl transferase activities (Haydock *et al.*, 1995). The derived consensus sequence motif identified by Haydock *et al.* for the methylmalonate transferase domain is **RVDVV-7-MxSxAXhW**, where h represents an aliphatic hydrophobic amino acid, X is either arginine, serine, alanine or glutamic acid and those residues found in all AT domains are omitted. The consensus sequence proposed for the malonyl transferase domain is **ETGYA-7-QxAxFGLL**. As can be seen in the sequence alignments shown in figure 5.6, the fungal MSAS sequences are reasonably similar to the malonyl transferase consensus, particularly in the latter half of this region, although they also match the methylmalonyl transferase

motif in several places. In contrast, the corresponding sequence in PKS2, RTDVV-7-HLGIAAVL, is a much better fit to the methylmalonyl consensus motif and is less similar to the malonyl consensus motif. This suggests the possibility that PKS2 may possess a methylmalonyl transferase activity, but as this would be unprecedented for a fungal PKS it seems more likely that the consensus motifs of Haydock *et al.* may not be usefully applicable to the fungal PKS enzymes.

In the multifunctional type I animal FAS, the acyltransferase domain is also known to be responsible for translocation of the acetyl starter unit onto the acyl carrier protein. While this dual specificity for acetyl-CoA and malonyl-CoA must have a structural basis in the acyltransferase domain, this has not yet been elucidated (Rangan & Smith, 1997).

The conserved amino acid motif **vfvFPQGaQW** that bounds the acyltransferase domain at the N-terminus (Donadio & Katz, 1992) can be clearly identified in PKS2 between residues 586-595. The C-terminal end of the domain, identified by the conserved motif **Gvavdwxxa** in the EryA polypeptides (Donadio & Katz, 1992) is positioned in the PKS2 sequence at residue 906, resulting in a domain of 321 amino acids. Domain lengths for the EryA acyltransferase domains have been variously reported as 345-350 amino acids (Donadio & Katz, 1992) and 414-430 amino acids (Bevitt *et al.*,1992).

5.3.2.3 Dehydratase domain

A conserved sequence element of **HxxxGxxxxP** has been identified around the active site histidine residue in several PKS and FAS enzymes (Bevitt *et al.*, 1992; Donadio & Katz, 1992; Joshi & Smith, 1993). The histidine residue at the beginning of this motif has been shown to have an important role in catalysis of the dehydration reaction (Joshi & Smith, 1993). As can be seen from the alignment shown in figure 5.6, this motif is clearly conserved in PKS2 as in the other "MSAS-type" PKS enzymes, although the Glycine residue is not universally conserved among PKSs. The region around the active site histidine is highly conserved between *P. patulum* PKS2 and other "MSAS-type" fungal PKS systems (figure 5.6). Although the boundaries of this domain have not been precisely defined to date, the N-terminus in MSAS appears to be located close to the boundary of the acyltransferase domain with no intervening linker region (Bevitt *et al.*, 1992). The extent of homology between PKS enzymes in this region suggests that the dehydratase domain is relatively short (Bevitt *et al.*, 1992; Donadio & Katz, 1992), extending for 140 to 170

amino acids to its C-terminus, which is consistent with the size of the type II *E. coli* FabA dehydratase (170 amino acids).

5.3.2.4 Ketoreductase domain

The ketoreductase domain of PKS2 possesses two conserved functional motifs, both of which have been identified in many PKS and FAS enzymes. The sequence **GxGxxG,** which is believed to bring about a turn between the first strand of a ß-sheet and the start of the following α-helix (Wierenga *et al.*, 1985), indicates the position of the NADPH binding site. This feature has been identified in many PKS and FAS enzymes and also in a slightly modified form in NADPH-dependant dehydrogenases (Scrutton *et al.*, 1990). The ketoreductase domain active site can be identified by its homology to the **FSSxxxxxG** motif (Cortes *et al.*, 1990, figure 5.6). The consensus sequence **PxGTvLv** identified in EryA (Donadio & Katz, 1992) which defines the beginning of this domain can be identified in PKS2 at residues 1416-1422. As yet, a C-terminal consensus sequence for this domain has not been defined. Donadio and Katz (1992) suggested that this domain extends for about 190 amino acids from the NADPH binding site, resulting in a size of approximately 200 amino acids in PKS2; however the size attributed to the MSAS ketoreductase domain by Bevitt *et al.* (1992) is some 50 residues larger than this and this latter size fits in better with the extent of similarity between MSAS and PKS2 (figure 5.7).

5.3.2.5 Acyl carrier protein

The acyl carrier protein is responsible for binding the growing polyketide backbone, via a thioester linkage to its phosphopantetheine arm. This domain can be clearly identified at the C-terminal end of the predicted *PKS2* gene product by homology to other "MSAS-type" PKS enzymes. The putative phosphopantetheine binding active site residue, Ser_{1741}, is located within a sequence motif that differs from the established consensus **LGxDS** (Cortes *et al.*, 1990) at the first residue, which is methionine in PKS2. This is a fairly conservative substitution, of hydrophobic amino acids, which is also apparent in *A. parasiticus* PKS1 (Walsh, 1993). Consensus sequences defining the boundaries of this domain have not as yet been identified, although the N- and C-termini in MSAS of *P. patulum* have been assigned by alignment with the Ery2 ACP domains (Bevitt *et al.*, 1992). Alignment of PKS2 with the three other "MSAS-type" predicted gene products suggests an N-terminus

for its ACP domain at Val_{1717} and a C-terminus at Phe_{1777}, resulting in a domain size of 61 amino acids.

5.3.2.6 No enoylreductase or thioesterase domain can be identified in the predicted *PKS2* gene product

The continuous homology to MSAS (figure 5.2) suggests that the predicted PKS2 polypeptide will possess the same functional domains, viz. the five discussed above, and will lack enoylreductase and thioesterase catalytic sites. To confirm that these functional domains were not present in PKS2 the amino acid sequence was examined for the presence of signature motifs for these active sites in regions of unassigned function. As the linear order of enzymic domains is conserved between the various type I FAS and PKS enzymes characterised to date, the region of predicted *PKS2* gene product on the C-terminal side of the dehydratase domain would be the best candidate for an enoylreductase domain; however no close matches to the enoylreductase NADPH binding site conserved motif **LIHxxxGGxGxA** (Yang *et al.*, 1996) could be found within this region. Hence the polyketide product of PKS2 must show incomplete processing of its keto groups, resulting in the possible presence of keto, hydroxyl and enoyl, but not alkyl, functional groups on the carbon chain.

Some fungal PKS enzymes, such as *A. nidulans* pksST and *A. parasiticus* pksL1 are known to possess a serine-dependant thioesterase domain which is thought to be responsible for the release of the completed PKS molecule from the enzyme (Feng & Leonard, 1995; Yu & Leonard, 1995). In contrast, no such thioesterase domain has been identified in the *MSAS* gene product and this enzyme is believed to employ a different mechanism for product release (see section 1.3.3.2.1). PKS2 shows almost continuous homology to MSAS along its length and contains only one region, between the predicted dehydratase and ketoreductase domains (figure 5.7), of sufficient size to contain a thioesterase domain and which is not already assigned an enzymic function. No motif resembling the consensus for the thioesterase active site (Feng & Leonard, 1995) could be identified within this region nor in any reading frame after the predicted stop codon (figure 5.1, nucleotides 5517-5519).

Figure 5.6 Alignment of the regions around the active sites of *P. patulum* PKS2 with other PKS and FAS enzymes.

The KS acyl-binding cysteine, AT pantetheine-binding serine, DH active histidine and ACP phosphopantetheine binding serine are indicated with asterisks, along with conserved residues in the NADPH binding and active site motifs of the ketoreductase domain. The acyltransferase sequences are shown aligned with the consensus motifs proposed by Haydock *et al.* (1995) beginning at the first residue.

Sequence identity codes are: PpPKS2, *P. patulum* PKS2; PpMSAS, *P. patulum* MSAS; AtMSAS, *A. terreus* MSAS; ApPKS1, *A. parasiticus* PKS1; ApPKSL1, *A. parasiticus* PKSL1; ChPKS1, *Cochliobolus heterostrophus* PKS1; ClPKS1, *Colletotrichum lagenarium* PKS1; AnwA, *A. nidulans* wA; RatFAS, *Rattus norvegicus* FAS; RmNODF, *Rhizobium meliloti* NODF; SeEry1/2/3, *Saccharopolyspora erythraea* Ery1/2/3; EcFabD, *E. coli* FabD; MbMAS, *Mycobacterium bovis* MAS.

Ketosynthase

```
                                    *
 PpPKS2   RISYLLDLMGPSVALDAACASSLVAVHHARQAIRA
 AtMSAS   RISYHLNLMGPSTAVDAACASSLVAIHHGRQAILQ
 PpMSAS   RISYHLNLMGPSTAVDAACASSLVAIHHGVQAIRL
 ApPKS1   RISYHLNLMGPSTAVDAACASSLVAIHLGRQAILS
 ClPKS1   RINYHFGFSGPSLNVDTACSSSAAALNVACNSLWQ
   AnwA   RINYYFKFSGPSVSVDTACSSSLAAIHLACNSIWR
 AnStcA   RINFCFEFSGPSYSNDTACSSSLAAIHLACNSLWR
ApPKSL1   RINFCFEFAGPSYTNDTACSSSLAAIHLACNSLWR
 ChPKS1   RISYSFDLKGPSVLVDTACSGGLTALHLACQSLLV
 SeERY2   RISYTMGLEGPSISVDTACSSSLVALHLAVESLRK
 RatFAS   RLSFFFDFKGPSIALDTACSSSLLALQNAYQAIRS
 ScActI   EVAWAVGAEGPVTMVSTGCTSGLDSVGNAVRAIEE
```

Acyltransferase

```
                                          *
 PpPKS2   RTDVVQAMTFLMHLGIAAVL.EAESGPPTAVVGHSLGEAAAAVVSGALTWHE
 AtMSAS   SSIKVQVLTYLVQVGLAAIL.RSKGLEPQAVIGHSVGEIAASVAAGCLTAEE
 ApPKS1   SSDQVQVLTYLMQIGLSEVL.RSLGVSCGAVIGHSVGEIAASVAAGCITPAE
 PpMSAS   SSDRVQILTYVMQIGLSALL.QSNGITPQAVIGHSVGEIAASVVAGALSPAE
ApPKSL1   APVVVQLAITCLQMALTN.LMTSFGIRPDVTVGHSLGEFAALYAAGVLSASD
 ClPKS1   SPVIVQLGLCCFEMALAR.LWASWGIRPSAVMGHSLGEYAALNAAGVLSASD
 ChPKS1   KALYSQPLCTALQIALVDLLV.SWGIYAQSVTGHSSGEIAAAYAAGALSLSD
   AnwA   SPVVVQLGTTCVQMALSS.FWASLGITPSFVLGHSLGDFAAMNAAGVLSTSD
 RatFAS   IVHSFFVSLTAIQIALIDLLT.SMGLKPDGIIGHSLGEVACGYADGCLSQRE
  MbMAS   GIDKVQPAVFAVQVALAATMEQTYGVRPGAVVGHSMGESAAAVVAGALSLED
 SeERY3   RVDVVQPALFAVMVSLAA.LWRSHGVEPAAVVGHSQGEIAAAHVAGALTLED
 EcFABD   KTWQTQPALLTASVALYRVWQQQGGKAPAMMAGHSLGEYSALVCAGVIDFAD
```

Dehydratase

```
                          *
PpPKS2   RLEEDNRPFPGRHPLHGSEIVPAAVLLHTFL
ApPKS1   RVDEQTKPFPGNHPVHGCEIIPAAVLMNTFL
PpMSAS   RLDNDTKPFPGSHPLHGTEIVPAAGLINTFL
AtMSAS   QMDDQTKPFPGSHPLHGSEIVPAAALVNTFL
SeERY2   RLSTDEQPWLAEHVVGGRTLVPGSVLVDLAL
RatFAS   ASSESSDHYLVDHCIDGRVLFPGTGYLYLVW
 MbMAS   DVGTSVLSWLSDHQVHNVAALPGAAYCEMAL
ChPKS1   VIRLSELPWLQDHKIQSSILYPVAGYIAMAI
```

Ketoreductase: NADPH binding site

```
                  *  *   *
PpPKS2   TYLITGGLGCLGLSVAEWMVTQ.GAKRILLLSRR
AtMSAS   TYLITGGLGALGLEVAQFLVEK.GARRLILVSRR
PpMSAS   TYLITGGLGVLGLEVADFLVEK.GARRLLLISRR
ApPKS1   TYLITGGLGDLGLATADFLVQK.GARRIILVSRR
ChPKS1   SYLLVGGVGGLGSATALWMSTR.GARHLLLLNRS
RatFAS   SYIITGGLGGFGLELARWLVLR.GAQRLVLTSRS
 MbMAS   SYIITGGLGGLGLFFASKLAAA.GCGRIVLTARS
SeERY3   TVLVTGGTGGIGAHVARWL.ARSGAEHLVLLGRR
SeERY2   TVLITGGTGTLGRLLARHLVTEHGVRHLLLVSRR
```

Ketoreductase: active site

```
                    ***       *
PpPKS2   LDFFILFSSCGQLLGFPGQASYASGNSFLD
AtMSAS   LDFFMLFSSCGQLLGFPGQASYASGNAFLD
ApPKS1   VDFMVLFSSCGQFFGFPGQASYASGNAFLD
PpMSAS   VDFFVMFSSCGNLVGFTGQASYGSGNAFLD
SeERY3   LDAFVLFSSNAGVWGSPGLASYAAANAFLD
ChPKS1   LDFFILESSISGIIGNPGQAAYAAANTFLD
 MbMAS   LDWFCLFSSGAALLGSPGQGAYAAANSWVD
SeERY1   LTAFVLFSSFASAFGAPGLGGYAPGNAYLD
RatFAS   LDYFVAFSSVSCGRGNAGQSNYGFANSTME
```

Acyl Carrier Protein

```
                      *
 PpPKS2   VDETVALPEMGMDSVMTVNFRMSLQQTL
 ApPKS1   VASQEPLSNMGMDSVMTVHLRGRLQKSL
 AtMSAS   VDSKAALSDLGVDSVMTVSLRGQLQKTL
 PpMSAS   VDSKAALADLGVDSVMTVTLRRQLQLTL
ClPKS1a   LADNIAFTDLGCDSLMALTVSGRMREEL
  AnwA2   MSDDLVFADYGVDSLLSLTVTGKYREEL
ClPKS1b   IIAAPDLAALGMDSLMSLSILGTLREKS
  AnwA1   IKSDENLNELGMDSLLSLTVLGKIRESL
 ChPKS1   VHTTKSLQDYGIDSLVAVELRNWLIKDM
 RatFAS   INLDSSLADLGLDSLMGVEVRQILEREH
 RmNODF   LTAATELTALGVDSLGLADIIWDVEQAY
```

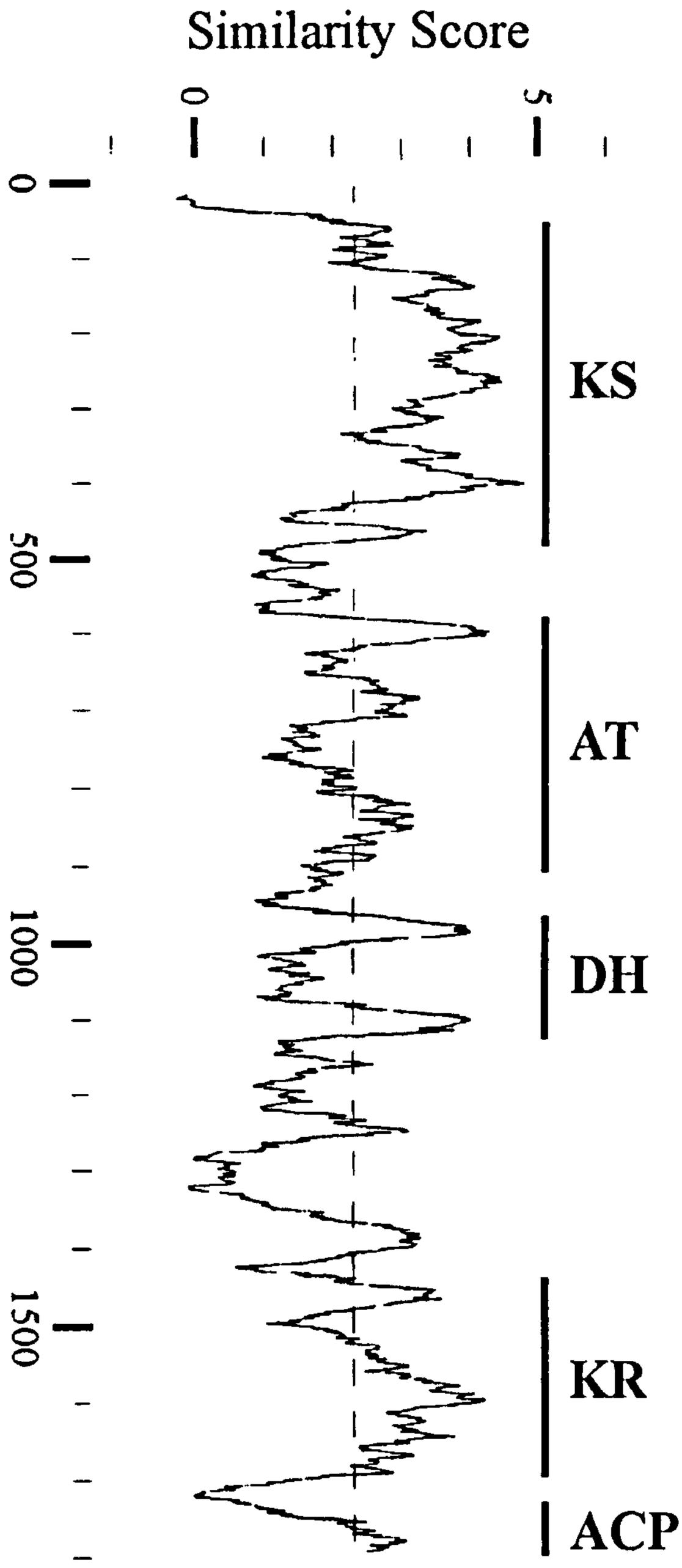


Figure 5.7 Similarity plot for *P. patulum* PKS2 and MSAS.

The two protein sequences were aligned using the program GAP. Similarity between the two aligned sequences was measured and plotted by the program PLOTSIMILARITY using a window size of 30. The predicted extents of the five functional domains are shown by the solid bars to the right of the plot. Average similarity across the entire alignment is plotted as a dotted line.

5.3.3 Interdomain regions

As can be seen from the domain structure shown in figure 5.7, there are relatively short interdomain regions situated between ketosynthase-acyltransferase, acyltransferase-dehydratase and ketoreductase-ACP domains. Bevitt *et al.* (1992) speculated that similar interdomain regions in EryA function either as rigid connectors, keeping domains apart, or as flexible linkers between enzymic domains which are important in facilitating catalysis and transfer of the growing polyketide between active sites. In the animal FAS these poorly-conserved linkers have been shown to be susceptible to proteolysis and are therefore believed to be exposed at the surface of the folded protein (Witkowski *et al.*, 1991). A much larger region of undefined function is located between the dehydratase and ketoreductase domains, and a similar extensive interdomain region is present in many other type I PKS and FAS between DH and ketoreductase/enoylreductase domains. The region is highly conserved in some modular PKS enzymes and is also resistant to limited proteolysis (Aparicio *et al.*, 1994; Aparicio *et al.*, 1996; Donadio & Katz, 1992), indicating that it has a specific functional significance, as yet unknown. A similar long interdomain region in the animal FAS is believed to located in the core of the functional dimer formed by this enzyme and is involved in stabilising subunit contacts (Joshi & Smith, 1993; Witkowski *et al.*, 1991).

5.3.4 Possible polyketide products for PKS2

6-methylsalicylic acid is an important precursor in the biosynthesis of fungal metabolites and is known to be involved in the biosynthesis of *P. patulum* metabolites other than those in the patulin pathway (figure 3.1). Examples include 6-formylsalicylic acid and 3-hydroxyphthalic acid which are formed by successive oxidations at the methyl group of 6-methylsalicylic acid (Bassett & Tanenbaum, 1958). While these metabolites may simply be by-products of the patulin biosynthetic pathway, it is possible that their biosynthesis involves a dedicated functional homologue of MSAS encoded by a separate gene which has diverged in sequence. *Penicillium patulum* is also known to produce the antibiotic griseofulvin along with several other heptaketide metabolites: fulvic acid, norlichexanthone and griseoxanthones C and B, which are all believed to be by-products of griseofulvin biosynthesis (Turner & Aldridge, 1983). These heptaketide metabolites are all believed to be derived from the benzophenone intermediate shown in figure 5.8. The carbon chain of

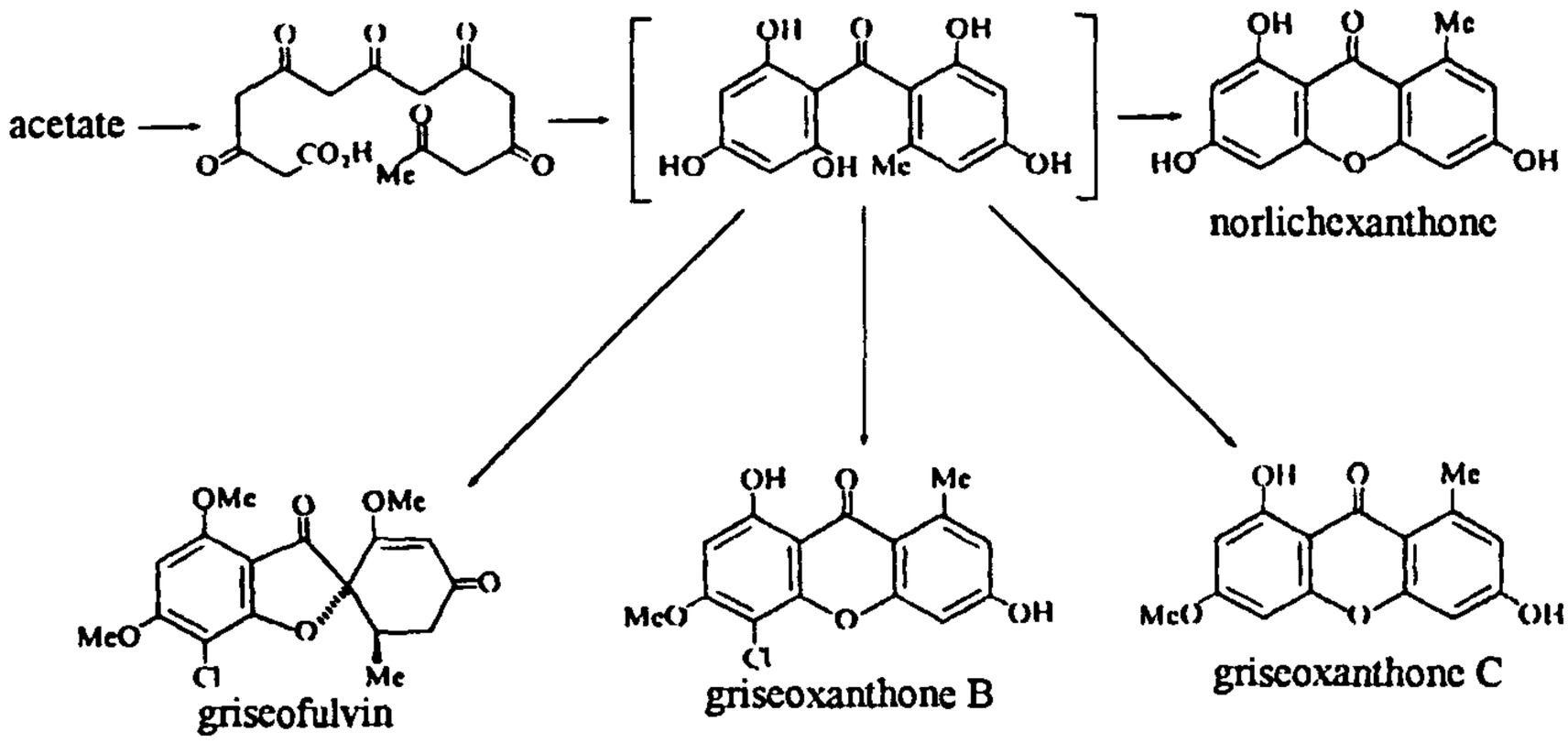


Figure 5.8 Biosynthesis of heptaketide metabolites of *Penicillium patulum*.

this parental compound has keto, hydroxy and enoyl functional groups consistent with the ketoreductase and dehydratase catalytic sites of PKS2. Optimal culture conditions for griseofulvin and patulin production by *P. patulum* are known to be different (Jimenez *et al.*, 1990); if this is due to differences in transcriptional regulation of the respective biosynthetic pathway genes, it could explain why *PKS2* does not appear to be expressed under the same conditions as *MSAS* (figures 5.4, 5.5).

Lastly *PKS2* may represent an inactive copy of a gene (or pseudogene) which in related strains of *Penicillium* is involved in the biosynthesis of polyketide metabolites not produced by *P. patulum* IMI 92273 (the strain used for this study). The 2.4 kb transcript that hybridises to the *PKS2* ketosynthase domain probe (figure 5.4) may be a truncated (and hence inactive) transcript arising from *PKS2*. Alternatively this band may represent hybridisation to the transcript of another, shorter, gene the product of which possesses a structurally related ketosynthase domain. Non-expressed homologues of aflatoxin biosynthesis genes have been identified in non-aflatoxigenic species of *Aspergillus* (Geisen, 1996; Klich *et al.*, 1995; Watson *et al.*, 1995).

5.3.5 A cytochrome p450 monooxygenase enzyme gene is closely linked to *PKS2* on the chromosome.

Similarity searches of the SwissProt database indicated the presence of a cytochrome p450 gene located immediately downstream from *PKS2* and transcribed in the opposite direction (figure 5.3). The two regions of homology to cytochrome p450s are encoded by different reading frames on the complementary strand; along with the positions of these regions on the chromosome this suggests that there is an intron located between them. A heme binding site (see section 3.3.6) and the conserved **ExxR** motif (Yu *et al.*, 1997) can be identified in the predicted protein product. As previously discussed (section 1.5.2), cytochrome p450 enzymes are often involved in the pathways of secondary metabolism and genes involved in the same secondary metabolic pathway tend to be clustered on the chromosome in filamentous fungi. This closely linked gene could well play a role in the processing of the polyketide product of PKS2.

CHAPTER 6

CONCLUSION

6.1 Putative patulin pathway genes

The experiments described in chapter 3 were directed towards the identification and characterisation of genes involved in the patulin biosynthetic pathway. Linkage to *MSAS* and a strong similarity to cytochrome P450 genes make *ORF2* a good candidate for a pathway gene, although there is no direct evidence for such a role. Cytochrome P450 enzymes are often involved in the biosynthesis of natural products, in addition to industrial biotransformations such as steroid compound production (Ohkawa *et al.*, 1990). The identification and characterisation of a putative novel fungal cytochrome P450 gene is of additional interest because information on these genes is limited (Hohn *et al.*, 1995). As prolific producers of natural products, the filamentous fungi may be assumed to possess a large number of these enzymes, as they are often involved in secondary metabolism. An *E. coli* expression vector was used in an attempt to assay the enzymatic activity of ORF2 protein as this was the simplest system available. This system has previously been used to produce functional fungal cytochrome P450 enzymes as discussed in section 3.3.9. Preliminary results from these assays have failed to identify any activity corresponding to a patulin pathway enzyme in crude cell extracts. This may mean that this enzyme is not involved in patulin biosynthesis, alternatively the protein expressed in *E. coli* may be inactive. The main use of *E. coli* expression systems has been for the expression of soluble, prokaryotic forms of cytochrome P450 and it is possible that expression of an authentic, functional ORF2 protein will require the use of a system more closely related to *P. patulum.* Yeast (*Saccharomyces cerevisiae*) and baculovirus expression vectors are a well-established means for the expression of functional eukaryotic cytochrome P450 enzymes (Gonzalez *et al.*, 1991; Gonzalez & Korzekwa, 1995; Guengerich *et al.*, 1991). In this laboratory both expression systems are currently in use (Lazarus & Macdonald, 1996; Macdonald *et al.*, 1994) and could be used if the protein expressed in *E. coli* appears to be non-functional. Additional characterisation of the structure of *ORF2* may also be worthwhile, in particular to identify the beginning and end points of transcription.

This study focused on ORF2 as it was possible to make some predictions about the function of the protein. Characterisation of ORF1 is currently in progress, following a similar approach to that taken for ORF2, and *E.coli* clones expressing what appears to be a full length protein have been produced. The gene product will be subjected to a similar functional analysis as is

underway with ORF2, although there is no indication as yet from sequence analysis that the ORF1 gene encodes a pathway enzyme. Novel classes of enzyme are being added continuously to protein databases; regular checking of the predicted protein sequence of ORF1 against these databases may yet provide some clue as to the function of this polypeptide.

Patulin biosynthesis is one of the best understood pathways of secondary metabolism, so elucidation of the genetics and enzymology involved is of great interest. If the assay of enzymatic function for the protein encoded by *ORF2* should reveal that it is capable of catalysing a reaction in the patulin biosynthetic pathway, then this would be taken to confirm that the pathway genes are clustered around *MSAS* on the chromosome. Sequence and northern blot analysis of the second chromosomal fragment upstream of *MSAS,* obtained by the chromosome walk described in section 3.2.2, would be of immediate interest. Further chromosome walking may also be worthwhile, in order to assess the extent of a gene cluster. The lack of a good system for homologous integration of DNA into the chromosomes of *P. patulum* may represent a problem in terms of identification of open reading frames involved in patulin biosynthesis. As in this study, heterologous expression of functional protein may provide an indication of the role of such genes. The use of a high efficiency transformation system such as that described by Chakraborty *et al.* (1991) may provide a low level of homologous integration sufficient for these purposes. Finally the research group of Dr. G.M. Gaucher at the University of Calgary have identified and characterised several patulin pathway mutants (Gaucher *et al.*, 1981), which could be useful in complementation analysis.

6.2 Novel molecular tools for the study of PKS genes

The LC-series PCR primers discussed in Chapter 4 will certainly be of interest to researchers interested in elucidating the molecular genetics of polyketide melanin or pigment biosynthesis in filamentous fungi. It is not possible to predict the range of fungal PKS genes that these primers are likely to amplify, as knowledge of sequence divergence between fungal PKS ketosynthase domains is very limited. Nonetheless, the LC1/2c primer pair might provide homologous probes for the PKS gene involved in a particular biosynthetic pathway and thereby a means for cloning it. For instance, a PCR product was obtained from the genome of *Verticillium dahliae,* an economically important pathogen of crop plants such as cotton,

tomato and potato. This could be used as a probe to facilitate cloning the corresponding PKS gene, which is likely to be involved in melanin production. The usefulness of the LC3/5c primer pair really depends on whether they are amplifying from *MSAS* homologues in the species examined or from a range of PKS genes. In the latter instance they may provide a useful molecular tool for any researcher wishing to isolate a novel PKS gene. As this primer pair appears to amplify from *PKS2* and *MSAS* in *P. patulum* (section 4.3.2), a functional analysis of *PKS2* (see below) may indicate whether LC3/5c will be useful for the isolation of novel PKS genes.

The phylogenetic tree constructed using the LC1/2c PCR product sequences (figure 4.13) gives a limited amount of information about the evolutionary relationships between the fungal PKS genes. This is a consequence of the small number of protein sequences that are available for analysis, and could be improved upon if more of the PCR products obtained using the LC-series primers (section 4.3.1) were sequenced. The cladogram does broadly support the division of the fungal PKS genes into two 'subgroups', the basis for the design of the LC-series primers. Cloning and sequence analysis of the genes to which these primers bind would help to confirm the hypothesis that the two 'subgroups' of fungal PKS genes differ in their ability to process keto groups (section 4.4.4).

6.3 Characterisation of *PKS2*

Polyketide biosynthesis in the filamentous fungi is a subject worthy of study both as a model for secondary metabolism in general and also because many of the end products are of economic or medical importance. Any new fungal PKS gene is interesting both in terms of its product and also for the elucidation of structure-function relationships in the PKS enzyme. While a certain amount of useful information has been derived from the *PKS2* gene sequence discussed in chapter 5, this only takes on real scientific value once the product of this gene can be identified. The immediately striking feature of the PKS2 enzyme is its similarity to MSAS. If PKS2 is found to catalyse a significantly different set of reactions to MSAS, for example producing the benzophenone intermediate shown in figure 5.8, this would raise interesting questions about the structural basis of the mechanistic differences between the two enzymes. Further studies of *PKS2* expression should be carried out to see if the gene is transcribed under

different culture conditions from those used in this study, in particular under conditions that are optimal for griseofulvin production. As discussed above, a high-efficiency transformation system for *P. patulum* might allow the product of PKS2 to be identified by gene disruption. Alternatively the activity of the gene product could be assessed via expression in the *Streptomyces coelicolor* CH999 system (Bedford *et al.*, 1995), which is capable of expressing large fungal PKS enzymes in an active form.

References

Albertini AM, Caramori T., Scoffone F, Scotti C and Galizzi A (1995) Sequence around the 159 degree region of the *Bacillus subtilis* genome: the *pksX* locus spans 33.6 kb. *Microbiol.* **141:** 299-309.

Alberts AW and Greenspan MD (1984) Animal and bacterial fatty acid synthetase: structure, function and regulation. In: *Fatty acid metabolism and its regulation.* Ed. S. Numa. Elsevier, Amsterdam.

Anderson JA, Lin B-K and Wang SS (1990) Purification and properties of emodin deoxygenase from *Pyrenochaeta terrestris*. *Phytochemistry* **29:** 2415-2418.

Andersen JF, Tatsuta K, Gunji H, Ishiyama T and Hutchinson CR (1993) Substrate specificity of 6-deoxyerythronolide B hydroxylase, a bacterial cytochrome P450 of erythromycin A biosynthesis. *Biochemistry* **32:** 1905-1913.

Aparicio JF, Caffrey P, Marsden AFA, Staunton J and Leadlay PF (1994) Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase.*J. Biol. Chem.* **269:** 8524-8528.

Aparacio JF, Molnar I, Schwecke T, Konig A, Haydock SF, Khaw LE, Staunton J and Leadlay PF (1996) Organisation of the biosynthetic gene-cluster for rapamycin in *Streptomyces hygroscopicus* - analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169:** 9-16.

Aramayo R, Adams TH and Timberlake WE (1989) A large cluster of highly expressed genes is dispensible for growth and development in *Aspergillus nidulans*. *Genetics* **122:** 65-71.

Ayer WA and Jimenez LD (1994) Phomalone, an antifungalmetabolite of *Phoma etheridgei. Can. J. Chem.* **72:** 2326-2332.

Barnes HJ, Arlotto MP and Waterman MR (1991) Expression and enzymatic activity of recombinant cytochrome P450 17α-hydroxylase in *Escherichia coli. Proc. Natl. Acad. Sci.* **88:** 5587-5601.

Bassett EW and Tanenbaum SW (1958) The metabolic products of *Penicillium patulum* and their probable inter-relationship. *Experientia* **14:** 38-40.

Beck J, Ripka S, Siegner A, Schiltz E and Schweizer E (1990) The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*: Its gene structure relative to that of other polyketide synthases. *Eur. J. Biochem.* **192:** 487-498.

Bedford D, Jacobsen JR, Luo G, Cane DE and Khosla C (1996) A functional chimeric modular polyketide synthase generated via domain replacement. *Chem. Biol.* **3:** 827-831.

Bedford DJ, Schweizer E, Hopwood DA and Khosla C (1995) Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **177:** 4544-4548.

Bennett JW (1995) From molecular genetics and secondary metabolism to molecular metabolites and secondary genetics. *Can. J. Bot.* **73:** S917-S924.

Berse B, Dmochowska A, Skrzypek M, Weglenski P, Bates MA and Weiss RL (1983) Cloning and characterisation of the ornithine carbamoyl-transferase gene from *Aspergillus nidulans. Gene* **25:** 109-117.

Betina V (1995) Differentiation and secondary metabolism in some prokaryotes and fungi. *Folia Microbiol.* **40:** 51-67.

Betina V (1989) Patulin and other small lactones.In *Mycotoxins. Chemichal, biological and environmental aspects*. pp. 243-270. Ed. V Betina. Elsevier, Amsterdam.

Bevitt DJ, Cortes J, Haydock SF and Leadlay PF (1992) 6-deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea. Eur. J. Biochem.* **204:** 39-49.

Birch AJ (1967) Biosynthesis of polyketides and related compounds. *Science* **156:** 202-206.

Birch AJ and Donovan FW (1953) Studies in relation to biosynthesis I. Some possible routes to derivatives of orcinol and phloroglucinol. *Aust J. Chem.* **6:** 360-368.

Birch AJ, English RJ, Massy-Westropp RA, Slaytor M and Smith H (1958) Studies in relation to biosynthesis. Part XIV. The origin of the nuclear methyl groups in mycophenolic acid. *J. Chem. Soc.* 365-368.

Birch AJ, Massy-Westropp RA, Rickards RW and Smith H (1958) Studies in relation to biosynthesis. Part XIII. Griseofulvin. *J. Chem. Soc.* 360-365.

Blanco G, Brian P, Pereda A, Méndez C, Salas JA and Chater KF (1993) Hybridisation and DNA sequence analyses suggest an early evolutionary divergence of related biosynthetic gene sets encoding polyketide antibiotics and spore pigments in *Streptomyces* spp. *Gene* **130:** 107-116.

Bowyer P, Osbourn AE and Daniels MJ (1994) An "instant gene bank" method for heterologous gene cloning: complementation of two *Aspergillus nidulans* mutants with *Gaeumannomyces graminis* DNA. *Mol. Gen. Genet.* **242:** 448-454.

Brobst SW and Townsend CA (1994) The potential role of fatty acid initiation in the biosynthesis of the fungal aromatic polyketide aflatoxin B_1. *Can. J. Chem.* **72:** 200-207.

Brown DW, Hauser FM, Tomassi R and Corlett S (1993) Structural elucidation of a putative conidial pigment intermediate in *Aspergillus parasiticus*. *Tetrahedron Lett.* **34:** 419-422.

Brown DW and Salvo JJ (1994) Isolation and characterisation of sexual spore pigments from *Aspergillus nidulans*. *App. Env. Microbiol.* **60:** 979-983.

Brown DW, Yu J-H, Kelkar HS, Fernandes M, Nesbitt TC, Keller NP, Adams TH and Leonard TJ (1996) Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*. *Proc. Natl. Acad. Sci.* **93:** 1418-1422.

Brown JA, Alic M and Gold MH (1991) Manganese peroxidase gene transcription in *Phanerochaete chrysosporium*: activation by manganese. *J. Bacteriol.* **173:** 4101-4106.

Bu'Lock JD (1980) Mycotoxins as secondary metabolites. In *The biosynthesis of mycotoxins.* pp. 1-16. Ed. PS Steyn. Academic Press, New York.

Bu'Lock JD, Detroy RW, Hoštálek Z and Munim-Al-Shakarchi A (1974) Regulation of secondary biosynthesis in *Gibberella fujikuroi*. *Trans. Br. Mycol. Soc.* **62:** 377-389.

Cary JW, Wright M, Bhatnagar D, Lee R and Chu FS (1996) Molecular characterisation of an *Aspergillus parasiticus* dehydrogenase gene, *norA*, located on the aflatoxin biosynthesis gene cluster. *Appl. Environ. Microbiol.* **62:** 360-366.

Casadaban MJ and Cohen SN (1980) Analysis of gene control signals by DNA and fusion in *Escherichia coli*. *J. Mol. Biol.* **138:** 179-207.

Cavener DR and Ray SC (1991) Eukaryotic start and stop translation sites. *Nucl. Acids. Res.* **19:** 3185-3192.

Chain E, Florey HW, Jennings MA and Callow D (1942) An antibacterial substance produced by *Penicillium claviforme*. *Br. J. Exp. Pathol.* **23:** 202-205.

Chakraborty BN, Patterson NA and Kapoor M (1991) An electroporation-based system for high-efficiency transformation of germinated conidia of filamentous fungi. *Can. J. Microbiol.* **37:** 858-863.

Chang PK, Cary JW, Yu J, Bhatnagar D and Cleveland TE (1995) The *Aspergillus parasiticus* polyketide synthase gene *pksA*, a homolog of *Aspergillus nidulans wA*, is required for aflatoxin B1 biosynthesis. *Mol Gen. Genet.* **248:** 270-277.

Chang, PK, Cary, JW, Bhatnagar, D, Cleveland, TE, Bennett, JW, Linz, JE, Woloshuk, CP and Payne GA (1993) Cloning of the *Aspergillus parasiticus apa*-2 gene associated with the regulation of aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **59:** 3273-3279.

Chang PK, Skory CD and Linz JE (1992) Cloning of a gene associated with aflatoxin B1 biosynthesis in *Aspergillus parasiticus*. *Curr. Genet.* **21:** 231-233.

Chiang T-Y and Marzluf GA (1994) DNA recognition by the NIT2 nitrogen regulatory protein: importance of the number, spacing, and orientation of GATA core elements and flanking sequences upon NIT2 binding. *Biochemistry* **33:** 576-582.

Child CJ, Spencer JB, Bhogal P and Shoolingin-Jordan PM (1996) Structural studies between 6-methylsalicylic acid synthase from *Penicillium patulum* and vertebrate type I fatty acid synthase: evidence from thiol modification studies. *Biochemistry* **35**: 12267-12274.

Collie JN (1907) Derivatives of the multiple ketene group. *Proc. Chem. Soc.* **91:** 230.

Compton T (1990) Degenerate primers for DNA amplification. In: *PCR protocols: a guide to methods and applications*. pp. 46-53. Eds. Innis MA, Gelfand DH, Sninsky JJ and White TJ. Academic Press, London.

Cortes J, Haydock SH, Roberts GA, Bevitt DJ and Leadlay PF (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348:** 176-178.

Cortes J, Wiesmann KEH, Roberts GA, Brown MJB, Staunton J and Leadlay PF (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268:** 1487-1489.

Coupland K and Niehaus Jr. WG (1987) Effect of nitrogen supply, Zn^{2+} and salt concentration on kojic acid and versicolorin biosynthesis by *Aspergillus parasiticus*. *Exp. Mycol.* **11:** 206-213.

Cox RJ, Hitchman TS, Byrom KJ, Findlow ISC, Tanner JA, Crosby J and Simpson TJ (1997) Post-translational modification of heterologously expressed *Streptomyces* type II polyketide synthase acyl carrier proteins. *FEBS Lett.* **405:** 267-272.

Crosby J, Sherman DH, Bibb MJ, Revill WP, Hopwood DA and Simpson TJ (1995) Polyketide synthase acyl carrier proteins from streptomyces - expression in *Escherichia coli*, purification and partial characterisation. *Biochim. Biophys. Acta* **1251:** 32-42

Crump MP, Crosby J, Dempsey CE, Parkinson JA, Murray M, Hopwood DA and Simpson TJ (1997) Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **36:** 6000-6008.

Dalboge H and Heldthansen HP (1994) A novel method for efficient expression cloning of fungal enzyme genes. *Mol. Gen. Genet.* **243:** 253-260.

Davis NK and Chater KF (1990) Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol. Microbiol.* **4:** 1679-1691.

Dawson MJ, Farthing JE, Marshall PS, Middleton RF, O'Neill MJ, Shuttleworth A, Stylli C, Tait M, Taylor PM, Wildman HG, Buss AD, Langley D and Hayes MV (1992) The squalestatins, novel inhibitors of squalene synthase produced by a species of *Phoma*. *J. Antibiotics* **45:** 639-647.

Decker H, Motamedi H and Hutchinson CR (1993) Nucleotide sequences and heterologous expression of *tcmG* and *tcmP*, biosynthetic genes for tetracenomycin C in *Streptomyces glaucescens*. *J. Bacteriol.* **175:** 3876-3886.

Devereux J, Haeberli P and Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res.* **12:** 387-395.

Dimroth P, Ringelmann E and Lynen F (1976) 6-methylsalicylic acid synthetase from *Penicillium patulum*. Some catalytic properties of the enzyme and its relation to fatty acid synthetase. *Eur. J. Biochem.* **68:** 591-596.

Dimroth P, Walter H and Lynen F (1970) Biosynthese von 6-methylsalicylsaüre. *Eur. J. Biochem.* **13:** 98-110.

Dobinson KF (1995) Genetic transformation of the vascular wilt fungus *Verticillium dahliae*. *Can. J. Bot.* **73:** 710-715.

Donadio S and Hutchinson CR (1991) Cloning and characterisation of the *Saccharopolyspora erythraea fdxA* gene encoding ferredoxin. *Gene* **100:** 231-235.

Donadio S and Katz L (1992) Organisation of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111:** 51-60.

Donadio S, McAlpine JB, Sheldon PJ, Jackson M and Katz L (1993) An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc. Natl. Acad. Sci. USA* **90:** 7119-7123.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ and Katz L (1992) Biosynthesis of the erythromycin macrolactone and a rational approach to producing hybrid macrolides. *Gene* **115:** 97-103.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ and Katz L (1991) Modular organisation of genes required for complex polyketide biosynthesis. *Science* **252:** 675-679.

Dutton MF (1988) Enzymes and aflatoxin biosynthesis. *Microbiol. Rev.* **52:** 274-295.

Eckert KA and Kunkel TA (1994) The fidelity of DNA polymerases used in the polymerase chain reactions. In *PCR: a practical approach.* Eds. MJ McPherson, P Quirke and GR Taylor. IRL Press, Oxford.

Ehrenshaft M and Upchurch RG (1991) Isolation of light-enhanced cDNAs of *Cercospora kikuchii*. *Appl. Environ. Microbiol.* **57:** 2671-2676.

Ellis CM (1996) Penicillium urticae nitrogen regulatory factor NRFA gene. *EMBL database* **U53137.**

Epp JK, Huber MLB, Turner JR, Goodson T and Schoner BE (1989) Production of a hybrid macrolide antibiotic in *Streptomyces ambofaciens* and *Streptomyces lividans* by introduction of a cloned carbomycin biosynthetic gene from *Streptomyces thermotolerans*. *Gene* **85:** 293-301.

Feng GH, Chu FS and Leonard TJ (1992) Molecular cloning of genes related to aflatoxin biosynthesis by differential screening. *Appl. Environ. Microbiol.* **58:** 455-460.

Feng GH and Leonard TJ (1995) Characterisation of the polyketide synthase gene (*pksL1*) required for aflatoxin biosynthesis in *Aspergillus parasiticus*. *J. Bacteriol.* **177:** 6246-6254.

Fernandes ND and Kolattukudy PE (1996) Cloning, sequencing and characterisation of a fatty acid synthase-encoding gene from *Mycobacterium tuberculosis* var. *bovis* BCG. *Gene* **170:** 95-99.

Fernández-Moreno MA, Martínez E, Boto L, Hopwood DA and Malpartida F (1992) Nucleotide sequence and deduced functions of a set of co-transcribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem.* **267:** 19278-19290.

Fischer M, Durand R and Fevre M (1995) Characterization of the promoter region of the enolase encoding gene *enol* from the anaerobic fungus *Neocallimastix frontalis* - sequence and promoter analysis. *Curr. Genet.* **28:** 80-86

Friedman L (1990) Patulin: Mycotoxin or fungal metabolite? (Current state of knowledge). *Biodeterior. Res.* **3:** 21-54.

Fu H, Ebert-Khosla S, Hopwood DA and Khosla C (1994) Engineered biosynthesis of novel polyketides: dissection of the catalytic specificity of the *act* ketoreductase. *J. Am. Chem. Soc.* **116:** 4166-4170.

Fu YH, Marzluf GA (1990) *nit-2*, the major nitrogen regulatory gene of neurospora-crassa, encodes a protein with a putative zinc finger DNA-binding domain. *Mol. Cell. Biol.* **10:** 1056-1065.

Fujii I, Ebizuka Y and Sankawa U (1988) A novel anthraquinone ring cleavage enzyme from *Aspergillus terreus*. *J. Biochem.* **103:** 878-883.

Fujii I, Ono Y, Tada H, Gomi K, Ebizuka Y and Sankawa U (1996) Cloning of the polyketide synthase gene *atX* from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. *Mol. Gen. Genet.* **253:** 1-10.

Gallie DR and Bailey-Serres J (1997) Eyes off transcription! The wonderful world of post-transcriptional regulation. *Plant Cell* **9:** 667-673.

Gaucher GM, Lam KS, Grootwassink JWD, Neway J and Deo YM (1981) The initiation and longevity of patulin biosynthesis. *Developments in Industrial Microbiology* **22:** 219-232.

Geisen R (1996) Multiplex polymerase chain reaction for the detection of potential aflatoxin and sterigmatocystin producing fungi. *System. Appl. Microbiol.* **19:** 388-392.

Gems D, Aleksenko A, Belenky L, Robertson S, Ramsden M, Vinetski Y and Clutterbuck AJ (1994) An "instant gene bank" method for gene cloning by mutant complementation. *Mol. Gen. Genet.* **242:** 467-471.

Giles NH, Case ME, Baum J, Geever R, Huiet L, Patel V and Tyler B (1985) Gene organisation and regulation in the *qa* (quinic acid) gene cluster of *Neurospora crassa*. *Microbiol. Rev.* **49:** 338-358.

Gill M (1994) Pigments of fungi (macromycetes). *Nat. Prod. Rep.* **11:** 67-90.

Girgis SI, Alevizaki M, Denny P, Ferrier GJM and Legon S (1988) Generation of DNA probes for peptides with highly degenerate codons using mixed PCR. *Nucl. Acids Res.* **16:** 10371.

Gish W and States DJ (1993). Identification of protein coding regions by database similarity search. *Nature Genet.* **3:** 266-272.

Gonzalez FJ (1989) The molecular biology of cytochrome P450s. *Pharmacol. Rev.* **40:** 243-284.

Gonzalez FJ, Kimura S, Tamura S and Gelboin HV (1991) Expression of mammalian cytochrome-P450 using baculovirus. *Methods Enzymol.* **206:** 93-99

Gonzalez FJ and Korzekwa KR (1995) Cytochromes P450 Expression Systems. *Ann. Rev Pharmacol. Toxicol.* **35:** 369-390

Gould SJ, Subramani S and Scheffler IE (1989) Use of the DNA polymerase chain reaction for homology probing. Isolation of partial cDNA or genomic clones encoding the iron-sulfur protein of succinate dehydrogenase from several species. *Proc. Natl. Acad. Sci. USA* **86:** 1934-1938.

Grimm A, Madduri K, Ali A and Hutchinson CR (1994) Characterisation of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **151:** 1-10.

Grootwassink JWD and Gaucher GM (1980) De novo biosynthesis of secondary metabolism enzymes in homogeneous cultures of *Penicillium urticae*. *J. Bacteriol.* **141:** 443-455.

Guengerich FP, Brian WR, Sari M-A and Ross JT (1991) Expression of mammalian cytochrome P450 enzymes using yeast-based vectors. *Methods Enzymol.* **206:** 130-145.

Gurr SJ, Unkles SE and Kinghorn JR (1987) The structure and organisation of nuclear genes of filamentous fungi. In *Gene structure in eukaryotic microbes.* Ed. J.R. Kinghorn. IRL Press, Oxford.

Gurusiddaiah S and Ronald RC (1981) Grahamimycins: antibiotics from *Cytospora* sp. Ehrenb. W.F.P.L. 13A. *Antimicrob. Agents Chemother.* **19:** 153-165.

Gyllensten UB and Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. *Proc. Natl. Acad. Sci. USA* **85:** 7652-7656.

Haas H and Marzluf GA (1995) NRE, the major nitrogen regulatory protein of *Penicillium chrysogenum*, binds specifically to elements in the intergenic promoter regions of nitrate assimilation and penicillin biosynthetic gene clusters. *Curr. Genet.* **28:** 177-183.

Ha-Huy-Ke and Luckner M (1979) Structure and function of the conidiospore pigments of *Penicillium cyclopium. Z. Allg. Mikrobiol.* **19:** 117-122.

Hallam SE, Malpartida F and Hopwood DA (1988) Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor. Gene* **74:** 305-320.

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166:** 557-580.

Haslam E (1986) Secondary metabolism - fact and fiction. *Nat. Prod. Rep.* **3:** 217-249.

Haydock SF, Aparicio JF, Molnár I, Schwecke T, Khaw LE, König A, Marsden AFA, Galloway IS, Staunton J and Leadlay PF (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Letters* **374:** 246-248.

Herbert RB (1989) *The biosynthesis of secondary metabolites.* Chapman and Hall, London.

Hohn TM, McCormick SP and Desjardins AE (1993) Evidence for a gene cluster involving trichothecene-pathway biosynthetic genes in *Fusarium sporotrichioides. Curr. Genet.* **24:** 291-295.

Holzer KP, Liu W and Hammes GG (1989) Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA. *Proc. Natl. Acad. Sci. USA* **86:** 4387-4391.

Hopkins J (1993) The toxological hazards of patulin. *Fd Chem. Toxic.* **31:** 455-456.

Hopwood DA (1988) Towards an understanding of gene switching in *Streptomyces*, the basis of sporulation and antibiotic production. *Proc. R. Soc. Lond.* B **235:** 121-138.

Hopwood DA and Khosla C (1992) Genes for polyketide secondary metabolic pathways in microorganisms and plants. *Ciba Found. Symp.* **171:** 88-112.

Hopwood DA, Malpartida F, Kieser HM, Ikeda H, Duncan J, Fujii I, Rudd BAM, Floss HG and Omura S (1985) Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314:** 642-644.

Hopwood DA and Sherman DH (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu. Rev. Genet.* **24:** 37-66.

Horng, JS, Chang PK, Pestka JJ and Linz JE (1990) Development of a homologous transformation system for *Aspergillus parasiticus* with the gene encoding nitrate reductase. *Mol. Gen. Genet.* **224:** 294-296.

Huang K-X, Fujii I, Ebizuka Y, Gomi K and Sankawa U (1995) Molecular cloning and heterologous expression of the gene encoding dihydrogeodin oxidase, a multicopper blue enzyme from *Aspergillus terreus. J. Biol. Chem.* **270:** 21495-21502.

Hunter IS (1992) Function and evolution of secondary metabolites-no easy answers. *Trends Biotechnol.* **10:** 144-146.

Hutchinson CR and Fujii I (1995) Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* **49:** 201-238.

Jayakumar A, Tai MH, Huang WY, al-Feel W, Hsu M, Abu-Elheiga L, Chirala SS and Wakil SJ (1995) Human fatty acid synthase: properties and molecular cloning. *Proc. Natl. Acad. Sci. USA* **92:** 8695-8699.

Jimenez M, Mateo R, Querol A, Mateo JJ and Hernandez E (1990) Differentiation of *Penicillium griseofulvum* Dierckx isolates by enzyme assays and by patulin and griseofulvin analyses. *Appl. Environ. Microbiol.* **56:** 3718-3722.

Jordan PM and Spencer JB (1993) The biosynthesis of tetraketides: enzymology, mechanism and molecular programming. *Biochem. Soc. Trans.* **21:** 222-228.

Joshi AK and Smith S (1993) Construction, expression, and characterisation of a mutated animal fatty acid synthase deficient in the dehydratase function. *J. Biol. Chem.* **268:** 22508-22513.

Junghanns KT, Kneusel RE, Baumert A, Maier W, Gröger D and Matern U (1995) Molecular cloning and heterologous expression of acridone synthase from elicited *Ruta graveolens* L. cell suspension cultures. *Plant Mol. Biol.* **27:** 681-692.

Kao CM, Katz L and Khosla C (1994) Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265:** 509-512.

Kao CM, Luo G, Katz L, Cane DE and Khosla C (1995) Manipulation of a macrolide ring size by directed mutagenesis of a modular polyketide synthase. *J. Am. Chem. Soc.* **117:** 9105-9106.

Katz L and Donadio S (1993) Polyketide synthesis: prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* **47:** 875-912.

Kelkar HS, Skloss TW, Haw JF, Keller NP and Adams TH (1997) *Aspergillus nidulans stcL* encodes a putative cytochrome P-450 monooxygenase required for bisfuran desaturation during aflatoxin/sterigmatocystin biosynthesis. *J. Biol. Chem.* **272:** 1589-1594.

Keller NP and Adams TH (1995) Analysis of a mycotoxin gene cluster in *Aspergillus nidulans*. *SAAS Bull. Biochem. Biotechnol.* **8:** 14-21.

Keller NP, Cleveland TE and Bhatnagar D (1992) A molecular approach towards understanding aflatoxin production. In *Mycotoxins in ecological systems.* Eds. Bhatnagar D, Lillehoj EB and Arora DK. Marcel Dekker, New York.

Keller NP, Dischinger Jr HC, Bhatnagar D, Cleveland TE and Ullah AHJ (1993) Purification of a 40-kilodalton methyltransferase active in the aflatoxin biosynthetic pathway. *Appl. Environ. Microbiol.* **59:** 479-484.

Keller NP and Hohn TM (1997) Metabolic pathway gene clusters in filamentous fungi. *Fungal Genet. Biol.* **21:** 17-29.

Keller NP, Kantz NJ and Adams TH (1994) *Aspergillus nidulans verA* is required for production of the mycotoxin sterigmatocystin. *Appl. Environ. Microbiol.* **60:** 1444-1450.

Keller NP, Segner S, Bhatnagar D and Adams TH (1995) stcS, a putative P-450 monooxygenase, is required for the conversion of versicolorin A to sterigmatocystin in *Aspergillus nidulans*. *Appl. Environ. Microbiol.* **61:** 3628-3632.

Khosla C, Ebert-Khosla S and Hopwood DA (1992) Targetted gene replacements in a *Streptomyces* polyketide synthase gene cluster: role for the acyl carrier protein. *Mol. Microbiol.* **6:** 3237-3249.

Khosla C, McDaniel R, Ebert-Khosla S, Torres R, Sherman DH, Bibb MJ and Hopwood DA (1993) Genetic construction and functional analysis of hybrid polyketide synthases containing heterologous acyl carrier proteins. *J. Bacteriol.* **175:** 2197-2204.

Khosla C and Zawada RJX (1996) Generation of polyketide libraries via combinatorial biosynthesis. *Trends Biotechnol.* **14:** 335-341.

Kim ES, Hopwood DA and Sherman DH (1994) Analysis of type II polyketide ß-ketoacyl synthase specificity in *Streptomyces coelicolor* A3(2) by *trans* complementation of actinorhodin synthase mutants. *J. Bacteriol.* **176:** 1801-1804.

Kimura N and Tsuge T (1993) Gene cluster involved in melanin biosynthesis of the filamentous fungus *Alternaria alternata. J. Bacteriol.* **175:** 4427-4435.

Kinghorn JR and Hawkins AR (1982) Cloning and expression in *Escherichia coli* K-12 of the biosynthetic dehydroquinase function of the *arom* cluster gene from the eukaryote, *Aspergillus nidulans. Mol. Gen. Genet.* **186:** 142-152.

Kiuchi N, Naruse A, Yamamoto H and Sekiguchi J (1991) Transformation of *Penicillium urticae* with plasmids containing the hygromycin B resistance gene. *Agric. Biol. Chem.* **55:** 3053-3057.

Klich MA, Yu J, Chang PK, Mullaney EJ, Bhatnagar D and Cleveland TE (1995) Hybridization of genes involved in aflatoxin biosynthesis to DNA of aflatoxigenic and non-aflatoxigenic aspergilli. *Appl. Microbiol. Biotechnol.* **44:** 439-443.

Köttig H, Rottner G, Beck K-F, Schweizer M and Schweizer E (1991) the pentafunctional *FAS1* genes of *Saccharomyces cerevisiae* and *Yarrowia lipolytica* are co-linear and considerably longer than previously estimated. *Mol. Gen. Genet.* **226:** 310-314.

Kreuzaler F and Hahlbrook K (1975) Enzymic synthesis of an aromatic ring from acetate units. Partial purification and some properties of flavonone synthase from cell-suspension cultures of *Petroselinium hortense. Eur. J. Biochem.* **56:** 205-213.

Kubo Y, Nakamura H, Kobaysashi K, Okuno T and Furusawa I (1991) Cloning of a melanin biosynthetic gene essential for appressorial penetration of *Colletotrichum lagenarium. Mol. Plant Microbe Interact.* **4:** 440-445.

Kubo Y, Tsuda M, Furusawa I and Shishiyama J (1989) Genetic analysis of genes involved in melanin biosynthesis of *Cochliobolus miyabeanus. Exp. Mycol.* **13:** 77-84.

Kudla B, Caddick MX, Langdon T, Martinez-Rossi NM, Bennett CF, Sibley S, Davies RW and Arst HN (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus* nidulans. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. *EMBO J.* **9:** 1355-1364.

Kurosaki F, Itoh M, Kizawa Y and Nishi A (1993) Partial purification and characterisation of a polyketide biosynthetic enzyme, 6-hydroxymellein synthase, in elicitor-treated carrot cell extracts. *Arch. Biochem. Biophys.* **300:** 157-163.

Lambalot RH and Cane DE (1995) Overproduction and characterisation of the erythromycin C-12 hydroxylase, EryK. *Biochemistry* **34:** 1858-1866.

Lanz T, Tropf S, Marner F-J, Schröder J and Schröder G (1991) The role of cysteines in polyketide synthases. *J. Biol. Chem.* **266:** 9971-9976.

Lazarus CM and Macdonald H (1996) Characterisation of a strawberry gene for auxin-binding protein, and its expression in insect cells. *Plant Mol. Biol.* **31:** 267-277.

Macdonald H, Henderson J, Napier RM, Venis MA, Hawes C, Lazarus CM (1994) Authentic processing and targeting of active maize auxin-binding protein in the baculovirus expression system

Plant Physiol. **105:** 1049-1057.

Lee CC and Caskey CT (1990) cDNA cloning using degenerate primers. In: *PCR protocols: a guide to methods and applications.* pp. 46-53. Eds. Innis MA, Gelfand DH, Sninsky JJ and White TJ. Academic Press, London.

Lee CC, Wu X, Gibbs RA, Cook RG, Muzny DM and Caskey CT (1988) Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. *Science* **239:** 1288-1291.

Lewin B (1997) *Genes VI.* Oxford University Press.

Liang SH, Wu TS, Lee R, Chu FS and Linz JE (1997) Analysis of mechanisms regulating expression of the *ver-1* gene, involved in aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **63:** 1058-1065.

Light RJ and Vogel G (1975) 6-methylsalicylic acid decarboxylase. *Methods Enzymol.* **43:** 530-540.

Linton KJ, Cooper HN, Hunter IS and Leadlay PF (1994) An ABC transporter from *Streptomyces longisporoflavus* confers resistance to the polyether-ionophore antibiotic tetronasin. *Mol. Microbiol.* **11:** 777-785.

Linz JE & Pestka JJ (1992) Mycotoxins: Molecular strategies for control. In *Aspergillus: Biology and industrial applications.* Eds. JW Bennet & MA Klich. pp217-231. Butterworth-Heinemann, Boston, USA.

Luengo JM and Peñalva MA (1994) Penicillin biosynthesis. In *Aspergillus: 50 years on.* Eds. S.D. Martinelli and J.R. Kinghorn. pp620-623. Elsevier, Amsterdam.

MacNeil DJ, Occi JL, Gewain KM and MacNeil T (1994) Correlation of the avermectin polyketide synthase genes to the avermectin structure. *Ann. NY Acad. Sci. USA* **721:** 123-132.

Mahanti N, Bhatnagar D, Cary JW, Joubran J and Linz JE (1996) Structure and function of *fas-1A*, a gene encoding a putative fatty acid synthetase directly involved in aflatoxin biosynthesis in *Aspergillus parasiticus*. *Appl. Environ. Microbiol.* **62:** 191-195.

Malpartida F, Hallam SE, Kieser HM, Motamedi H, Hutchinson CR, Butler MJ, Sugden DA, Warren M, McKillop C, Bailey CR, Humphreys GO and Hopwood DA (1987) Homology between *Streptomyces* genes coding for synthesis of different polyketides used to clone antibiotic biosynthesis genes. *Nature* **325:** 818-821.

Masel AM, Irwin JAG and Manners JM (1993) DNA addition or deletion is associated with a major karyotype polymorphism in the fungal phytopathogen *Colletotrichum gloeosporiodes. Mol. Gen. Genet.* **237:** 73-80.

Mathews CK and van Holde KE (1990) *Biochemistry* Benjamin/Cummings, Redwood City.

Mathison L, Soliday C, Stepan T, Aldrich T and Rambosek J (1993) Cloning, characterisation, and use in strain improvement of the *Cephalosporium acremonium* gene *cefG* encoding acetyl transferase. *Curr. Genet.* **23:** 33-41.

Matsushima K, Ando Y, Hamasaki T and Yabe K (1994) Purification and characterisation of two versiconal hemiacetal acetate reductases involved in aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **60:** 2561-2567.

Mayfield MB, Godfrey BJ and Gold MH (1994) Characterisation of the *mnp2* gene encoding manganese peroxidase isozyme 2 from the basidiomycete *Phanerochaete chrysosporium. Gene* **142:** 231-235.

Mayorga ME and Timberlake WE (1992) The developmentally regulated *Aspergillus nidulans wA* gene encodes a polypeptide homologous to polyketide and fatty acid synthases. *Mol. Gen. Genet.* **235:** 205-212.

Mayorga ME and Timberlake WE (1990) Isolation and molecular characterisation of the *Aspergillus nidulans wA* gene. *Genetics* **126:** 73-79.

McCabe PC (1989) Assymetric polymerase chain reaction. *Amplifications* **3:** 1-3.

McCormick SP, Hohn TM and Desjardins AE (1996) isolation and characterisation of *Tri3*, a gene encoding 15-*O*-acetyltransferase from *Fusarium sporotrichioides. Appl. Environ. Microbiol.* **62:** 353-359.

McDaniel R, Ebert-Khosla S, Hopwood DA and Khosla C (1995) Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375:** 549-554.

McDaniel R, Ebert-Khosla S, Hopwood DA and Khosla C (1993) Engineered biosynthesis of novel polyketides. *Science* **262:** 1546-1550.

Mizutani, M, Ward E, Dimaio J, Ohta D, Ryals J and Sato R (1993) Molecular-cloning and sequencing of a cDNA-encoding mung bean cytochrome-P450 (P450c4h) possessing cinnamate 4-hydroxylase activity. *Biochem. Biophys. Res. Commun.* **190:** 875-880.

Mohamed AH, Chirala SS, Mody NH, Huang W-Y and Wakil SJ (1988) Primary structure of the multifunctional α subunit protein of yeast fatty acid synthase derived from *FAS2* gene sequence. *J. Biol. Chem.* **263:** 12315-12325.

Moreno MA, Ramos MDC, González A and Suárez G (1987) Effect of ultraviolet light irradiation on viability and aflatoxin production by *Aspergillus parasiticus. Can. J. Microbiol.* **33:** 927-929.

Murphy G and Lynen F (1975) Patulin biosynthesis: the metabolism of *m*-hydroxybenzyl alcohol and *m*-hydroxybenzaldehyde by particulate prepartions from *Penicillium patulum. Eur. J. Biochem.* **58:** 467-475.

Murphy G, Vogel G, Krippahl G and Lynen F (1974) Patulin biosynthesis: the role of mixed function oxidases in the hydroxylation of *m*-cresol. *Eur. J. Biochem.* **49:** 443-455.

Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K and Nebert DW (1993) The P450 superfamily - update on new sequences, gene-mapping, accession numbers, early trivial names of enzymes, and nomeclature. *DNA Cell Biol.* **12:** 1-51.

Niesbach-Klösgen U, Barzen E, Bernhardt J (1987) Chalcone synthase genes in plants: a tool to study evolutionary relationships. *Mol. Evol.* **26:** 213-225.

O'Hagan D (1991) *The polyketide metabolites*. Ellis Horwood, London.

O'Hagan D (1995) Biosynthesis of fatty acid and polyketide metabolites. *Nat. Prod. Rep.***12:** 1-32.

O'Neill JA, Simpson TJ and Willis CL (1993) Biosynthesis of colletodiol and related polyketide macrodiolides in *Cytospora* sp. ATCC 20502: synthesis and metabolism of advanced intermediates. *J. Chem. Soc., Chem. Commun.*: 738-740.

Ohkawa H, Yabusaki Y, Sakaki T, Murakami H and Shibata M (1990) Hydroxylation reactions by recombinant yeast cells expressing P450 monooxygenases. *Ann. N. Y. Acad. Sci. USA* **613:** 37-43.

Okamoto N, Tsuruta K, Imai Y, Tomura D and Shoun H (1997) Fungal P450nor: expression in *Escherichia coli* and site-directed mutagenesis at the putative distal region. *Arch. Biochem. Biophys.* **337:** 338-344.

Orkin SH (1992) GATA-binding transcription factors in haematopoietic cells. *Blood* **80:** 575-581.

Orvehed M, Haggblom P, Soderhall K (1988) Nitrogen inhibition of mycotoxin production by *Alternaria alternata. Appl. Environ. Microbiol.* **54:** 2361-2364.

Panaccione DG, Scott-Craig JS, Pocard JA and Walton JD (1992) A cyclic peptide synthetase gene required for pathogenicity of the fungus *Cochliobolus carbonum* on maize. *Proc. Natl. Acad. Sci. USA* **89:** 6590-6594.

Payne GA, Nystorm GJ, Bhatnagar D, Cleveland TE and Woloshuk CP (1993) Cloning of the *afl-2* gene involved in aflatoxin biosynthesis from *Aspergillus flavus. Appl. Environ. Microbiol.* **59:** 156-162.

Pazoutova S, Linka M, Storkova S and Schwab H (1995) *Aspergillus terreus* putative polyketide synthase (pks) gene, complete CDS. *EMBL database* **U31329**

Pearson WR and Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85:** 2444-2448.

Penfold CN, Bender CL and Turner JG (1996) Characterisation of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. *Gene* **183:** 167-173.

Pieper R, Luo G, Cane DE and Khosla C (1995a) Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. *Nature* **378:** 263-266.

Pieper R, Luo G, Cane DE and Khosla C (1995b) Remarkably broad substrate specificity of a modular polyketide synthase in a cell-free system. *J. Am. Chem. Soc.* **117:** 11373-11374.

Porter TD and Larson JR (1991) Expression of mammalian P450s in *Escherichia coli. Methods Enzymol.* **206:** 108-116.

Poulose AJ, Bonsall RF and Kolattukudy PE (1984) Specific modification of the condensation domain of fatty acid synthase and the determination of the primary structure of the modified active site peptides. *Arch. Biochem. Biophys.* **230:** 117-128.

Punt PJ and van den Hondel CAMJJ (1992) Analysis of transcription control sequences of fungal genes. In *Molecular signals in plant-microbe communications.* Ed. DPS Verma. pp 13-38. CRC press, Florida.

Raitt DC, Bradshaw RE and Pillar TM (1994) Cloning and characterisation of the cytochrome c gene of *Aspergillus nidulans. Mol. Gen. Genet.* **242:** 17-22.

Rangan VS and Smith S (1997) Alteration of the substrate specificity of the malonyl-CoA/acetyl-CoA:acyl carrier protein *S*-acyltransferase domain of the multifunctional fatty acid synthase by mutation of a single arginine residue. *J. Biol. Chem.* **272:** 11975-11978.

Revill WP, Bibb MJ and Hopwood DA (1995) Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. *J. Bacteriol.* **177:** 3946-3952.

Rittenberg D and Bloch K (1945) The utilisation of acetic acid for the synthesis of fatty acids. *J. Biol. Chem.* **154:** 417.

Rittenberg D and Bloch K (1944) The utilisation of AcOH for fatty acid synthesis. *J. Biol. Chem.* **154:** 311-312.

Robinson JA (1991) Polyketide synthase complexes: their structure and function in antibiotic biosynthesis. *Phil. Trans. R. Soc. Lond.* B **332:** 107-114.

Robison K and Smith DR (1994) *Mycobacterium leprae* cosmid L518. *EMBL Database* **U00023.**

Rollins MJ and Gaucher GM (1994) Ammonium repression of antibiotic and intracellular proteinase production in *Penicillium urticae. Appl. Microbiol. Biotechnol.* **41:** 447-455.

Roth J, Leroith D, Collier ES, Watkinson A and Lesniak MA (1986) The evolutionary origins of intercellular communication and the maginot lines of the mind. *Ann. N.Y. Acad. Sci.* **463:** 1-11.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbour Press, New York.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463-5467.

Schöppner A and Kindl H (1984) Purification and properties of a stilbene synthase from induced cell suspension cultures of peanut. *J. Biol Chem.* **259:** 6806-6811.

Schorr R, Mittag M, Müller G and Schweizer E (1994) Differential activities and intramolecular location of fatty acid synthase and 6-methylsalicylic acid synthase component enzymes.*J. Plant Physiol.* **143:** 407-415.

Schröder G, Brown JWS and Schröder J (1988) Molecular analysis of resveratrol synthase. cDNA, genomic clones and relationship with chalcone synthase*Eur. J. Biochem.* **172:** 161-169.

Schröder J, Tropf S, Raiber S, Leser J and Schröder G (1996) Plant-specific polyketides: the chalcone/stilbene synthase superfamily. *Royal Society of Chemistry, International interdisciplinary conference. Polyketides: chemistry, biochemistry and molecular genetics.* Lecture abstract I.11.

Schüz R, Heller W and Hahlbrock K (1983) Substrate specificity of chalcone synthase from *Petroselinum hortense. J. Biol. Chem.* **258:** 6730-6734.

Schwecke T, Aparicio JF, Molnar I, Konig A, Khaw LE, Haydock SF, Oliynyk M, Caffrey P, Cortes J, Lester JB, Bohm GA, Staunton J and Leadlay PF (1995) The biosynthetic gene-cluster for the polyketide immunosuppresant rapamycin. *Proc. Natl. Acad. Sci. USA* **92:** 7839-7843.

Schweizer M, Roberts LM, Höltke H-J, Takabayashi K, Höllerer E, Hoffmann B, Müller G, Köttig H and Schweizer E (1986) the pentafunctional FAS1 gene of yeast: its nucleotide sequence and order of the catalytic domains. *Mol. Gen. Genet.* **203:** 479-486.

Schweizer M, Takabayashi K, Laux T, Beck K-F and Schreglmann R (1989) Rat mammary gland fatty acid synthase: localisation of the constituent domains and two functional polyadenylation/termination signals in the cDNA. *Nucleic Acids Res.* **17:** 567-586.

Scott AI, Beadling LC, Georgopapadakou NH and Subbarayan CR (1974) Biosynthesis of polyketides. Purification and inhibition studies of 6-methylsalicylic acid synthase.*Bioorg. Chem.* **3:** 238-248.

Scott RE, Jones A and Gaucher GM (1986a) Manganese and antibiotic biosynthesis. III. The site of manganese control of patulin production in *Penicillium urticae. Can. J. Microbiol.* **32:** 273-279.

Scott RE, Jones A, Lam KS and Gaucher GM (1986b) Manganese and antibiotic biosynthesis. I. A specific manganese requirement for patulin production in *Penicillium urticae. Can. J. Microbiol.* **32:** 259-267.

Scott RE, Lam KS and Gaucher GM (1986) Stabilization and purification of the secondary metabolism specific enzyme, *m*-hydroxybenzylalcohol dehydrogenase. *Can. J. Microbiol.* **32:** 167-175.

Scotti C, Piatti M, Cuzzoni A, Perani P, Tognoni A, Grandi G, Galizzi A and Albertini AM (1993) A *Bacillus subtilis* ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* **130:** 65-71.

Scrutton NS, Berry A and Perham RN (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343:** 38-43.

Sekiguchi J and Gaucher GM (1977) Conidiogenesis and secondary metabolism in *Penicillium urticae. Appl. Environ. Microbiol.* **33:** 147-158.

Serre L, Verbree EC, Dauter Z, Stuitje AR and Derewenda ZS (1995) The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 15-Å resolution. *J. Biol. Chem.* **270:** 12961-12964.

Shafiee A and Hutchinson CR (1987) Macrolide antibiotic biosynthesis: isolation and properties of two forms of 6-deoxyerythronolide B hydroxylase from *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *Biochemistry* **26:** 6204-6210.

Shen B and Hutchinson CR (1993a) Tetracenomycin F1 monooxygenase: oxidation of a naphthacenone to a naphthacenequinone in the biosynthesis of tetracenomycin C in*Streptomyces glaucescens*. *Biochemistry* **32:** 6656-6663.

Shen B and Hutchinson CR (1993b) Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. *Science* **262:** 1535-1540.

Sherman DH, Kim E-S, Bibb MJ and Hopwood DA (1992) Functional replacement of genes for individual polyketide synthase components in *Streptomyces coelicolor* A3(2) by heterologous genes from a different polyketide pathway. *J. Bacteriol.* **174:** 6184-90.

Siedow JN, Rhoads DM, Ward GC and Levings III CS (1995) The relationship between the mitochondrial gene *T-urf13* and fungal pathotoxin sensitivity in maize. *Biochem. Biophys. Acta* **1271:** 235-240.

Simpson TJ (1995) Polyketide biosynthesis. *Chem. Ind.*

Simpson TJ (1991) The biosynthesis of polyketides. *Nat. Prod. Rep.* **8:** 573-602.

Simpson TJ (1985) The biosynthesis of polyketides. *Nat. Prod. Rep.* **2:** 321-347.

Skory CD, Chang PK, Cary J and Linz JE (1992) Isolation and characterisation of a gene from *Aspergillus parasiticus* associated with the conversion of versicolorin A to sterigmatocystin in aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **58:** 3527-3537.

Smith MW, Feng D-F and Doolittle RF (1992) Evolution by aquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* **17:** 489-493.

Smith DJ, Burnham MKR, Bull JH, Hodgson JE, Ward JM, Browne P, Brown J, Barton B, Earl AJ and Turner G (1990) β-Lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. *EMBO J.* **9:** 741-747.

Smith TF and Waterman MS (1981) Identification of common molecular subsequences. *J. Mol. Biol.* **147:** 195-197.

Sogin ML (1990) Amplification of ribosomal RNA genes for molecular evolution studies. In *PCR protocols: a guide to methods and applications.* Eds. MA Innis, Gelfand DH, JJ Sninsky and TJ White. Academic Press, London.

Spencer JB and Jordan PM (1992a) Purification and properties of 6-methylsalicylic acid synthase from *Penicillium patulum. Biochem. J.* **288:** 839-846.

Spencer JB and Jordan PM (1992b) Investigation of the mechanism and steric course of the reaction catalysed by 6-methylsalicylic acid synthase from *Penicillium patulum* using (R)-[1-^{13}C;2-^{2}H]- and (S)-[1-^{13}C;2-^{2}H]malonates. *Biochemistry* **31:** 9107-9116.

Steyn PS (1992) The biosynthesis of polyketide-derived mycotoxins. *JEPTO* **11:** 47-59.

Stone MJ and Williams DH (1992) On the evolution of functional secondary metabolites (natural products). *Mol. Microbiol.* **6:** 29-34.

Stoops JK, Wakil SJ (1981) Animal fatty-acid synthetase - a novel arrangement of the beta-ketoacyl synthetase sites comprising domains of the 2 subunits. *J. Biol. Chem.* **256:** 5128-5133.

Summers RG, Ali A, Shen B, Wessel WA and Hutchinson CR (1995) Malonyl-coenzyme A:acyl carrier protein acyltransferase of *Streptomyces glaucescens* - a possible link between fatty acid and polyketide biosynthesis. *Biochemistry* **34:** 9389-9402.

Summers RG, Wendt-Pienkowski E, Motamedi H and Hutchinson CR (1993) The *tcmVI* region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* encodes the tetracenomycin F1 monooxygenase, tetracenomycin F2 cyclase, and, most likely, a second cyclase. *J. Bacteriol.* **175:** 7571-7580.

Summers RG, Wendt-Pienkowski E, Motamedi H and Hutchinson CR (1992) Nucleotide sequence of the *tcmII-tcmIV* region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* and evidence that the *tcmN* gene encodes a multifunctional cyclase-dehydratase-*O*-methyl transferase. *J. Bacteriol.* **174:** 1810-1820.

Tanabe T, Noda M, Furutani Y, Takai T, Takahashi H, Tanaka K-I, Hirose T, Inayama S and Numa S (1984) Primary structure of ß subunit precursor of calf muscle acetylcholine receptor deduced from cDNA sequence. *Eur. J. Biochem.* **144:** 11-17.

Tanaka C, Kubo Y and Tsuda M (1991) Genetic analysis and chracterisation of *Cochliobolus heterostrophus* colour mutants. *Mycol. Res.* **95:** 49-56.

Takano Y, Kubo Y, Shimizu K, Mise K, Okuno T and Furusawa I (1995) Structural analysis of *PKS1*, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium. Mol. Gen Genet.* **249:** 162-167.

Thiele DJ (1992) Metal-regulated transcription in eukaryotes. *Nucleic Acids Res.* **20:** 1183-1191.

Töpfer R and Martini N (1994) Molecular cloning of cDNAs or genes encoding proteins involved in *de novo* fatty acid biosynthesis in plants. *J. Plant Physiol.* **143:** 416-425.

Trail F, Chang P-K, Cary J and Linz JE (1994) Structural and functional analysis of the *nor-1* gene involved in the biosynthesis of aflatoxins by *Aspergillus parasiticus. Appl. Environ. Microbiol.* **60:** 4078-4085.

Trail F, Mahanti N, Rarick M, Mehigh R, Liang S-H, Zhou R and Linz JE (1995) Physical and transcriptional map of an aflatoxin gene cluster in *Aspergillus parasiticus* and functional disruption of a gene involved early in the aflatoxin pathway. *Appl. Environ. Microbiol.* **61:** 2665-2673.

Tsay JT, Oh W, Larson TJ, Jackowski S and Rock CO (1992) Isolation and characterisation of the ß-ketoacyl-acyl carrier protein synthase III gene (fabH) from *Escherichia coli* K12. *J. Biol. Chem.* **267:** 6807-6814.

Turner G (1991) Strategies for cloning genes from filamentous fungi. In *Applied molecular genetics of fungi.* pp. 29-43. Eds. JF Peberdy, CE Caten, JE Ogden and JW Bennett. Cambridge University Press.

Turner G (1994) Genetics of penicillin biosynthesis in *Aspergillus nidulans*. In *The genus Aspergillus (FEMS symposium)*. pp. 197-207. Eds. KA Powell, A Renwick & JF Peberdy. Plenum Press, New york.

Turner WB (1971) *Fungal metabolites.* Academic Press, London.

Turner WB and Aldridge DC (1983) *Fungal metabolites II.* Academic Press, London.

Unkles SE (1992) Gene organisation in industrial filamentous fungi. In *Applied molecular genetics of filamentous fungi.* pp. 28-53. Eds. JR Kinghorn and G Turner. Blackie, London.

van den Hondel CAMJJ and Punt PJ (1991) Gene transfer systems and vector development for filamentous fungi. In: *Applied molecular genetics of fungi.* Eds. Peberdy JF, Caten CE, Ogden JE and Bennett JW. Cambridge University Press. pp. 1-28.

Verdoes JC, Punt PJ, Vanderberg P, Debets F, Stouthamer AH and van den Hondel CAMJJ (1994) Characterisation of an efficient gene cloning strategy for *Aspergillus niger* based on an autonomously replicating plasmid - cloning of the *nicB* gene of *Aspergillus niger. Gene* **146:** 159-165.

Vidal-Cros A, Viviani F, Labesse G, Boccara M and Gaudry M (1994) Polyhydroxynapthalene reductase involved in melanin biosynthesis in *Magnaporthe grisea.* Purification, cDNA cloning and sequencing. *Eur. J. Biochem.* **219:** 985-992.

Viera J and Messing J (1982) The pUC plasmids- an M13MP7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19:** 259-268.

Vining LC (1992) Secondary metabolism, inventive evolution and biochemichal diversity - a review. *Gene* **115:** 135-140.

Walsh M (1993) Cloning Fungal polyketide synthase genes. PhD Thesis, University of Bristol.

Wang I-K, Reeves C and Gaucher GM (1991) Isolation and sequencing of a genomic DNA clone containing the 3' terminus of the 6-methylsalicylic acid polyketide synthase gene of *Penicillium urticae*. *Can. J. Microbiol.* **37:** 86-95.

Watanabe CM, Wilson D, Linz JE and Townsend CA (1996) Demonstration of the catalytic roles and evidence for the physical association of type I fatty acid synthases and a polyketide synthase in the biosynthesis of aflatoxin B1. *Chem. Biol.* **3:** 463-469.

Watson AJ, Archer DB, Seal S and Linz JE (1995) Homologues of aflatoxin biosynthetic genes in *Aspergillus oryzae*. *Fungal Genetics Newsletter* **42A:** 82.

Weber JM, Leung JO, Swanson SJ, Idler KB and McAlpine JB (1991) An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science* **252:** 114-117.

Wheeler MH (1983) Comparisons of fungal melanin biosynthesis in ascomycetous, imperfect and basidiomycetous fungi. *Trans. Br. mycol. Soc.* **81:** 29-36.

Wheeler MH and Klich MA (1995) The effects of tricyclazole, pyroquilon, phthalide, and related fungicides on the production of conidial wall pigments by *Penicillium* and *Aspergillus* species. *Pestic. Biochem. Physiol.* **52:** 125-136.

Wierenga RK, Demaeyer MCH, Hol WGJ (1985) Interaction of pyrophosphate moieties with alpha-helixes in dinucleotide binding-proteins. *Biochemistry* **24:** 1346-1357.

Wiesner P, Beck J, Beck K-F, Ripka S, Muller G, Lücke S and Schweizer E (1988) Isolation and sequence analysis of the fatty acid synthetase *FAS2* gene from *Penicillium patulum*. *Eur. J. Biochem.* **177:** 69-79.

Williams DH, Stone MJ, Hauck PR and Rahman SK (1989) Why are secondary metabolites (natural products) biosynthesised? *J. Nat. Prods.* **52:** 1189-1208.

Witkowski A, Rangan VS, Randhawa ZI, Amy CM and Smith S (1991) Structural organisation of the multifunctional animal fatty-acid synthase. *Eur. J. Biochem.* **198:** 571-579.

Wolff RA and Hull R (1996) A rapid and easy method for DNA recovery from agarose gels using Wizrd minicolumns. *Trends Genet.* **12:** 339-340.

Woloshuk CP, Seip ER, Payne GA and Adkins CR (1989) Genetic transformation system for the aflatoxin-producing fungus *Aspergillus flavus*. *Appl. Environ. Microbiol.* **55:** 86-90.

Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim M and Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res.* **15:** 529-541.

Yabe K, Ando Y, Hashimoto J and Hamasaki T (1989) Two distinct *O*-methyltransferases in aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **55:** 2172-2177.

Yang G, Rose MS, Turgeon BG and Yoder OC (1996) A polyketide synthase is required for fungal virulence and production of the polyketide T-toxin. *Plant Cell* **8:** 2139-2150.

Ye J, Dickens ML, Plater R, Li Y, Lawrence J and Strohl WR (1994) Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing *Streptomyces* sp. strain C5. *J. Bacteriol.* **176:** 6270-6280.

Yu J, Cary JW, Bhatnagar D, Cleveland TE, Keller N and Chu FS (1993) Cloning and characterisation of a cDNA from *Aspergillus parasiticus* encoding an *O*-methyltransferase involved in aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **59:** 3564-3571.

Yu J, Chang PK, Cary JW, Bhatnagar D and Cleveland TE (1997) *avnA*, a gene encoding a cytochrome P450 monooxygenase, is involved in the conversion of averantin to averufin in aflatoxin biosynthesis in*Aspergillus parasiticus*. *Appl. Environ. Microbiol.* **63:** 1349-1356.

Yu J, Chang P-K, Cary JW, Wright M, Bhatnagar D, Cleveland TE, Payne GA and Linz JE (1995) Comparative mapping of aflatoxin pathway gene clusters in*Aspergillus parasiticus* and *Aspergillus flavus*. *App. Environ. Microbiol.* **61:** 2365-2371.

Yu J, Chang P-K, Payne GA, Cary JW, Bhatnagar D and Cleveland TE (1995) Comparison of the *omtA* genes encoding *O*-methyltransferases involved in aflatoxin biosynthesis from*Aspergillus parasiticus* and *A. flavus*. *Gene* **163:** 121-125.

Yu JH and Leonard TJ (1995) Sterigmatocystin biosynthesis in*Aspergillus nidulans* requires a novel type I polyketide synthase. *J. Bacteriol.* **177:** 4792-4800.

Zotchev SB and Hutchinson CR (1995) Cloning and heterologous expression of the genes encoding nonspecific electron transport components for a cytochrome P450 system of*Saccharopolyspora erythraea* involved in erythromycin production. *Gene* **156:** 101-106.

ABBREVIATIONS

6MSA	6-methylsalicylic acid
6MSAS	6-methylsalicylic acid synthase
A	adenosine
ACP	acyl carrier protein
bp	base pair
C	centigrade
C	cytosine
cDNA	complementary deoxyribonucleic acid
CHS	chalcone synthase
CLF	chain length factor
CoA	coenzyme A
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DHN	dihydroxynapthalene
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetra-acetic acid
EtBr	ethidium bromide
FAS	fatty acid synthase
g	standard acceleration of gravity
G	guanosine
kb	kilobase
IPTG	isopropyl-1-thio-ß-D-galactoside
LCFA	long chain fatty acid
min	minute
mRNA	messenger ribonucleic acid
MRE	metal response element
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PKS	polyketide synthase
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RS	resveratrol synthase
RT	reverse transcription
SAM	*S*-adenosyl-L-methionine
SDS	sodium dodecylsulphate
SSC	standard saline citrate
ST	sterigmatocystin
STS	stilbene synthase
T	thymine
TAE	Tris-acetate buffer
TBE	Tris-borate buffer
TE	Tris-EDTA
TE	thioesterase
Tris	2-amino-2-hydroxymethylpropane-1,3-diol
U	uracil
UV	ultra violet